194-00549TPW

ECS Scientist User Survey (ESUS)

Technical Paper

Technical Paper—Not intended for formal review or government approval.

October 1994

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Stanley A. West /s/ 10/28/94

Stanley A. West, Senior Information Systems Engineer Date
EOSDIS Core System Project

SUBMITTED BY

Joy Beier /s/ 10/28/94

Joy Beier, Science Office Manager Date
EOSDIS Core System Project

Hughes Applied Information Systems
Landover, Maryland

This page intentionally left blank.

Abstract

ESUS (the ECS Scientist User Survey) was e-mailed to 434 EOS-funded scientists for their views about the shape that ECS should take and their anticipated modes of using ECS. The survey was administered electronically via INTERNET with an option to use Mosaic. Responses were received from 160 of the scientists (37 percent response rate). This paper describes the history of the ESUS Project, the response data, quantitative and qualitative data analyses, potential applications of the results to ECS development, and potential further work.

Responses to the survey questions illustrate the diversity of user opinions, with only a few items showing real consensus. The questions which produced consensus indicate the following user preferences:

- High quality and long-term data consistency are more important than having the most current data or the products from the latest algorithm updates;
- ECS should provide Level 3 and 4 data on a standard set of grids; support for user-specifiable grids was much lower, with a number of respondents indicating that this would be dangerous;
- ECS should provide tools for translating among different data formats, should provide some tools for data visualization, and should be able to accommodate scientist-provided tools.

The responses show little desire for multi-media capabilities within ECS, such as support for video conferencing, or video/audio annotation of data. However, even on these limited items, there were no nearly unanimous opinions.

Cluster analysis of the responses showed that 88 percent of the respondents (scientists) can be grouped into one of five groups (communities or subcultures) that have similar responses. The views of each of the five groups are detailed in Section 4.4.2 of this paper. Approximately 60 percent of the respondents belong to the two largest groups. Although ECS science users are diverse, much of that diversity can be captured by entering into dialog with typical representatives of each of only five groups of scientists as identified by ESUS.

The ESUS results show such great differences of opinion among the respondents that ECS should take care to avoid thinking that ECS has a singular science user community and constantly recognize that ECS has multiple science user communities.

This page intentionally left blank.

Contents

5. Potential Applications of the Results to ECS Development

6. Further Work

Figures

Tables

Appendix A. E-Mail Version of Questionnaire

Abbreviations and Acronyms

This page intentionally left blank.

1. Introduction

1.1 Purpose

As a part of the System Requirements Review (SRR) the ECS project was challenged to improve its awareness of and accommodation to the heterogeneity in the ECS's population of potential science users. ECS needs a process for dealing with the multiple perspectives that its science users will manifest. ECS needs a means to identify the typical users who can represent each of those multiple perspectives.

In order to rise to this challenge ECS must acquire an understanding of the scientists' diverse views, identify the scientists who espouse each view, identify typical scientists who subscribe to each major view, and encourage the participation of typical scientists in the shaping of ECS. This paper reports on the ESUS Project and its use of questionnaire methodology as a means to probe scientists' views about ECS. The questionnaire's character was shaped by a commitment to generate ECS-relevant data while minimizing the time required for scientists to complete the survey.

The present paper is intended to inform NASA and scientist customers and ECS staff members about the history and results of the ESUS Project. In the introduction to the questionnaire the ESUS Project committed to provide a summary of the survey results to all respondents who checked that they would like to receive results. This paper will be provided to all such interested respondents and be made Internet available using the ECS Data Handling System with Mosaic.

The potential applications of the ESUS results include user modeling, generally making ECS team members aware of the diversity of ECS science users, and formation of multiple electronic forums or special interest groups of like-minded scientists.

1.2 Organization

This paper is organized as follows:

Section 1 provides the purpose and general organization of this paper and delineates the procedures for its review and approval.

Section 2 chronicles the history of the development of the survey questionnaire.

Section 3 briefly describes the administration of the questionnaire.

Section 4 presents a summary of both quantitative and qualitative results.

Section 5 describes some potential applications of the results. Most potential applications will require enabling decisions by ECS or NASA management.

Section 6 identifies further work that could extend the ESUS Project to better understand the heterogeneity among the ECS science users.

1.3 Review and Approval

This Technical Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

The ideas expressed in this Technical Paper are valid for six months from the approval date.

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems
1616A McCormick Dr.
Landover, MD 20785

2. Development of the Survey Questionnaire Instrument

2.1 Science-based System Architecture Drivers for the ECS Project

While developing the questionnaire the ESUS Project made extensive use of the ECS White Paper "Science-based System Architecture Drivers for the ECS Project" (Revision 1.0, December, 1993). That paper summarized key architectural issues that emanated from the System Requirements Review. For example, Driver Number 6 consisted of "Support user-to-user collaboration" and was further elaborated by explanatory text and quoted comments by scientists. These design drivers were influential in forming the foundation for the conceptual architecture, which then formed a subsequent foundation for the SDR architecture. The ESUS Project converted the design drivers into language that is suitable for use in a questionnaire, added other issues as appropriate, enlisted ECS engineers and scientists for extensive reviewing of the language, and repeatedly revised the language without any intention to maintain fidelity to the drivers paper. That drivers paper began the development process, but reviewers' comments were used to direct the change in the evolving questionnaire according to the reviewers perceptions of ECS issues.

2.2 Revision and Expansion of Drivers into Pithy Statements

Effective questionnaires about technically-advanced subjects, such as ECS's desirable capabilities, are challenging to develop because each question should be succinct, present only one idea, and be understandable without further explanation. The original drivers had to be revised because the summary statements are too terse to be understood without explanation and because the explanations of the drivers discuss multiple ideas in a page or more of text. In its initial step the ESUS Project revised each of the drivers as one or more positively-phrased and simple statements.

For example, Science User Driver Number 1 is "facilitate an efficient data search and `access' paradigm." The White Paper devotes several paragraphs to explain that driver including the following. "The `search and order' paradigm is potentially too heavyweight and too bureaucratic. A lighter weight `search and order' paradigm should be employed, in which, once objects have been identified through specification in a search operation, they can simply be accessed (i.e., passed to an application, `opened,' etc.)"

We decomposed this general access issue into two sub-issues -- one about system access and another about data access. We used questions 2.23 and 2.24 (Appendix A) to probe the system access issue and questions 2.25 and 2.26 to probe the data access issue. As some comments showed, these issues presented considerable challenge to craft language that clearly communicates the intended ideas and packages the ideas as compact and simple statements that each communicate only one notion.

2.3 Conversion of the Statements into the Questionnaire Instrument

A list of statements derived from design drivers falls far short of a questionnaire. We built the questionnaire by transforming each statement into a question (Section 2 in the questionnaire). Each respondent is asked to agree or disagree with the statement by using a 5-point Likert scale. New text was written to introduce the survey, explain the background of the survey, provide instructions for responding to each statement, define the response scale, and return the responses to ECS. Questions about the respondent's background were also added (Section 1 in the questionnaire) in order to allow for the possibility that respondents' backgrounds might explain their attitudes as reflected by their responses in questionnaire Section 2.

Sample survey questionnaires usually are administered by mail or by telephone but not by multiple media. The ESUS instrument is doubly unusual because the questionnaire is administered electronically and has 2 forms. Respondents can choose to respond by E-Mail using ASCII text or by the graphical user interface provided by Mosaic.

2.3.1 E-Mail Version

All respondents initially receive the E-Mail version of the survey (Appendix A) as an Internet message. Respondents then choose between the E-Mail and Mosaic options as explained in the questionnaire. If they respond by E-Mail, respondents use an ASCII editor in order to insert their answers to the Internet message and then reply to us. For convenience in extracting comments from respondents' messages the ESUS Project requested respondents to type their comments at the end of the questionnaire. Not all respondents already use Mosaic. Therefore, the E-Mail version provided instructions about procedures to obtain Mosaic.

2.3.2 Mosaic Version

The Mosaic version of the questionnaire reproduces all questions that are in the E-Mail version but omits the instructions for obtaining Mosaic. The Mosaic version also differs by giving respondents the convenience of commenting in a comment box near each question rather than grouping all comments at the end of the questionnaire. Although Mosaic software is available for PC, UNIX, and Macintosh computers, the Macintosh Mosaic did not yet support the "forms" capability, and could not be used for the survey. The questionnaire was available to PC and Unix users.

2.4 Internal ECS Reviews

Internal ECS reviews were conducted in order to refine the questions, solicit contributions of additional questions that address ECS developers' concerns, and eliminate questions that address non-essential issues. The reviewers included the ESUS team members, ECS staff members in the Landover facility who will apply the results to system design, and the ECS DAAC scientists. Many revisions were made as the result of these reviews and every word was discussed repeatedly and at length.

2.5 Pretest by DAAC Scientists and the EOSDIS Project Scientist

In the pretest the ECS DAAC scientists were again recruited with the addition of the EOSDIS Project Scientist. Pretesters were asked to answer the questionnaire in order to anticipate weaknesses in the survey instrument rather than to generate data. A final series of revisions was stimulated by feedback such as the following. The draft had asked about willingness to wait for quality data without specifying the length of the waiting time. After a scientist said to "give some idea of the wait time," we elaborated the question as three questions that specify waits of "an hour," "a day," and "a week" (questions 2.16-2.18). Another scientist suggested that respondents should be encouraged to "check all that apply" in the background questions in questionnaire Section 1. The results proved the wisdom of this suggestion when many respondents made multiple checks for these background questions that show that scientists play multiple roles. A scientist correctly noted that we were asking about support of a standard set of grids for data products when Levels 1 and 2 are normally not gridded. We revised the questions 2.1 and 2.2 to focus only on Levels 3 and 4 although the comments in Section 4.2 of this paper show that details about levels confused several respondents. The DAAC scientists further supported ESUS by testing whether or not the questionnaire worked without computer errors on their local computers. Our greatest concern was with the Mosaic version.

This page intentionally left blank.

3. Administration of Questionnaire

3.1 E-Mail Broadcasts to EOS-Funded Scientists

Most surveys randomly sample a small subset of a large population of potential respondents such as the millions of households in the United States that have listed telephone numbers. Such samples are carefully crafted to avoid the cost of surveying the total population while producing conclusions that are representative of the total population. Pollsters and other survey researchers select a sample size that allows them to estimate statistical parameters, such as the true population means, with an acceptable confidence level that the sample parameters are within a tolerable error range from the true population parameters. For example, they ask “what sample size N is required in order to estimate the mean of a 5 point Likert scale subject to the requirement for a 95 percent confidence that the sample mean will be within 1 point of the true population mean?”

ESUS took another approach that has been termed opportunity sampling -- broadcast of questionnaires to all readily-identified Scientists. We sought scientists' names and Internet addresses in the most recent EOS Science Directory (1992) that was available in Spring 1994 and then retained principal and co-investigators, team leaders, and team members. This filtering was intended to eliminate non-practicing scientists such as governmental administrators. Although the resulting list is a subset of the potential ECS science users, that list includes nearly all of the funded scientists. Because ESUS elicits responses from all identified scientists, the data analysis focuses on surfacing the characteristics of the surveyed users rather than estimating the statistical parameters of some larger population with some specified confidence level. Section 6 identifies subsequent research that includes expansion of ESUS to identify and survey additional scientists but does not propose a random sampling design. ESUS was able to contact all identified scientists because large Internet surveys are nearly as economic to administer as small Internet surveys. The cost of the labor-intensive telephone and in-person surveys is proportion to the number of interviews so sampling all of a large population may be prohibitively expensive. The ESUS approach is consistent with the practice in the statistical field of "exploratory data analysis."

Beginning on May 26, 1994 we sent questionnaires to the EOS scientists. More than one attempt had to be made to reach some scientists because of outdated information in the Directory. Whenever a message was returned as undeliverable, we used Mosaic to access the NASA Gopher and conduct a search for the scientists who could not be reached. Using the Internet addresses that the Gopher search uncovered, we sent the questionnaire to all corrected addresses. We sent 434 questionnaires, including 25 that remained non-deliverable, and received 117 completed responses from this first phase of surveying. We have no way to know how many of the deliverable questionnaires reached the mail boxes of people who rarely or never check their mail.

3.2 Reminder by ECS Deputy Project Scientist

On June 6 the ECS Deputy Project Scientist sent a motivational Internet message to all non-respondents. This reminder stimulated 43 additional scientists to respond in this second phase of the survey and increased the total response to 160 (a 37 percent response rate). This 37 percent response rate compares favorably to the 25 to 30 percent response rate that mailed (not E-Mailed) questionnaires typically attain. Accumulated experience with E-Mail and Mosaic response rates is elusive because these media for questionnaire administration are still novel compared to telephone and mail. The responses to the present questionnaire included 113 E-Mail responses (71 percent) and 47 Mosaic responses (29 percent). The high rate of response by E-Mail does not necessarily imply a preference for E-Mail. Scientists may have been discouraged to reply by Mosaic if they did not already use Mosaic. Macintosh users would have been unable to use Mosaic (see Section 2.3.2).

4. Results of Questionnaire

4.1 Introduction

The main goal of the questionnaire was to systematically question a sizable number of scientists. Therefore, the multiple choice questions were emphasized in hope that all respondents would answer all questions. The multiple choice questions are crucial in that they support systematic comparisons among all or many respondents. When compared to the multiple choice responses, the comments are less valuable for systematic comparisons. The comments are essential as a rich source of insight about all the subjects that ESUS inadvertently overlooked. Although the questionnaire did not probe the level of scientist interest in opening a dialog between ECS and science users, one reasonable interpretation of this volume of comments would be that the scientists desire to enter into dialog with ECS about the future capabilities that will be available to users.

An ad hoc measurement concerns the time that scientists needed to complete the questionnaire using Mosaic. One-third of the Mosaic respondents needed 10 minutes or less. This suggests that we succeeded in fielding a questionnaire that can quickly be answered by busy scientists. These estimates may include time when the respondent was not actively completing the questionnaire and literally may include coffee or lunch breaks. All Mosaic respondents averaged 25 minutes to take the questionnaire with a 4 minute to two hour time range. We suspect that durations longer than an hour indicate absences, and after culling out such long durations the average falls to 17 minutes. The 17 minute estimate may also include periods when the respondent was distracted from the questionnaire. Comparable data are unavailable for E-Mail respondents because the returned questionnaires only show answers.

4.2 Comments on the Questionnaire Instrument

This section focuses on the comments that shed light on the validity or invalidity of the questionnaire. Respondents generally provided comments in order to identify problems that they perceive in the language used or other aspects of the questionnaire design although two praised Mosaic as a survey medium. Table 4-1 groups the comments by question numbers or topics that crosscut individual questions. Some comments have been paraphrased or summarized while making every effort to preserve the intent of the complete comment. The remainder of this section analyzes the comments and assesses the validity of the survey instrument.

These comments need to be considered in the context of the survey's goals. The character of the questionnaire was shaped by a commitment to generate ECS-relevant data while minimizing the time for scientists to complete the survey. We reasoned that scientists would be encouraged to respond to a questionnaire that takes only 10 minutes although we would have liked to probe each respondent for several times as much data. We resisted the temptation to try to accomplish every possible goal with one questionnaire when additional issues can always be probed by subsequent questionnaires. While developing the instrument we felt a continual tension among the guidelines for effective questionnaire design, a desire to generate richly detailed data, and a temptation to make the questionnaire show our understanding of the full complexity of the EOSDIS Core System. Showing unnecessary detail can harm a questionnaire by intimidating respondents who are unfamiliar with that information. After resolving to design a questionnaire that is quick to take we were obliged to omit much complexity in the explanatory text and questions.

Table 4-1. Comments on the Questionnaire Instrument (1 of 3)

Question or Topic of Comment	Comment
Question 1.11	It is unclear what kind of tool may be of general interest.
Question 2.1	Need to define Levels 3 and 4. (6 comments)
Question 2.2	Need to define Levels 3 and 4. (3 comments)
Question 2.3	Clarify if format is logical or physical.
Question 2.13	What is "video annotation"?
Question 2.16	Answer would depend on nature of the request.
Question 2.17	Make question more specific. Answer would depend on nature of the request.
Question 2.18	Say "wait ONLY 1 week" not "wait 1 week." Make question more specific. Answer would depend on nature of the request.
Question 2.19	Make question more specific. Answer would depend on nature of the request.
Question 2.20	Don't feel that there is one answer. It depends.
Question 2.23	2.23 and 2.24 are phrased to exclude each other. I would like to have automatic access but would be prepared to order it.
Question 2.24	Do not understand the difference between "ordering" and "specifying what I want." (2 comments)
Questions 2.23, 2.24, etc.	A number of investigators are doing analysis across a range of spatial and temporal scales, and it seems difficult to specify the most efficient way to get the data. It may vary with the data set and the role of that data within the analysis.
Question 2.25	Unclear or ambiguous (2 comments)

Table 4-1. Comments on the Questionnaire Instrument (2 of 3)

Question or Topic of Comment	Comment
General	Surveys like this help but are not enough. You also need to observe scientists for 2 weeks. This survey format is fantastic. Great way to collect a survey! Prefer to have a "neutral" response in addition to "no opinion." FYI - survey took 10 minutes to complete. Many of the questions are difficult to answer because they sound like motherhood and apple pie. (2 comments) Please note that this questionnaire needs sharpening up. I showed the survey to the DAAC manager and he was curious why it was being limited to the 550 EOS investigators. Clearly, the system is supposed to serve more than this group.
Cost Issues	Many of these questions have cost implications. It is not reasonable to ask such questions without providing some idea of the cost involved. I am bothered by this survey ... There are clear economic and bandwidth issues ... These should be discussed. Does this indicate where you are in ECS development? I should think that these sorts of profiles would have been developed 2 years ago.

Table 4-1. Comments on the Questionnaire Instrument (3 of 3)

Question or Topic of Comment	Comment
"It all depends"	These are hard questions with no right answer and a lot of the answers depend on specific uses and specific data sets -- good luck! Some questions numbered 2.1 to 2.26 are worded in such a way that agreeing with one question forces you to disagree with another question. (Note: intent of comment appears to be that one might agree with both questions under different circumstances.) Maybe I'm typical, but my research concerns phenomena on different time and spatial scales. So for me, this questionnaire was rather poorly devised. (Subsequent note by commenter: the problem is the assumption that each investigator will have one problem or one method of studying it that could easily be related to problem need. I think that you will find that most scientists are studying problems that require multiple temporal, spectral, and spatial scales of investigation.) Many of the questions are difficult to generalize since the answer depends completely upon the application. Some of these questions depend on what task is being accomplished. Many of us play different parts in EOS. Some of your questions are too simplistic, e.g., 2.23 and 2.24. In some cases I will want one, but in other, both need to be available.

The most careful designers of questionnaires expect some questions to be misunderstood and for some respondents to criticize every questionnaire. The present questionnaire appears to have succeeded well because only 1 comment made the global statement that "this questionnaire needs sharpening up" rather than having the many criticisms that were expected. Another comment suggested that such a survey should have been conducted 2 years ago. The following analysis of comments uncovers themes common to more than one respondent or question and also probes weaknesses in particular questions.

- Theme 1: “It all depends.”

Theme 1 appears in Table 4-1 in the row labeled by "it all depends" and in comments to questions 2.16, 2.17, 2.18, 2.19, 2.20, 2.23, and 2.24.

Scientists may respond differently to questions depending on the details of the situation that they picture as ECS science users. For example, a particular respondent might be both a data product developer and a data product consumer. Some answers could depend on whether that respondent focuses on the consumer role or the developer role. The data for the background questions in Section 1 of the questionnaire establish that most respondents identify with many of the background traits. This suggests that most respondents are complex users who will play many

roles when exploiting ECS as a tool for conducting science research. "It all depends" on the particular role or application that the respondent envisions while responding to the questions. As one commenter stated "These are hard questions with no right answer and a lot of the answers depend on specific uses and specific data sets -- good luck!" It is possible that the 2 criticisms that "many of the questions are difficult to answer because they sound like motherhood and apple pie" are actually reactions to the problem that "it all depends."

Subsequent questionnaires probably should be more tightly focused and specify situations more fully. The ESUS team made the judgment that the questionnaire would have become heavily burdened by adding language to exactly specify all of the details that surround each question. Specifying all assumed situations would have lengthened the questionnaire or have forced us to address fewer issues about ECS. We recognize that it does "all depend" and stand by the decision to omit all those dependencies in this exploratory questionnaire.

- Theme 2: Costs should be considered

Two commenters observed that the questionnaire omits the tradeoffs between ECS capabilities and the costs of providing those capabilities. This is another case of "it all depends" in the sense that the attractiveness of each capability will be influenced by the costs of all capabilities and the total resources that are available. One cannot conduct benefit-cost analyses without knowing costs! Cost data were not available to ESUS during the questionnaire design phase and inclusion of such information would have required a much more complex design if cost data were available.

- Theme 3: Why not survey all scientists?

This comment applies to this methodology section only in the sense that it proposes an expansion of the sample. Discussion of that issue will be deferred to Section 6.

- Theme 4: Prefer to have a "neutral" response in addition to "no opinion."

We acknowledge that the Mosaic version could have added a "no response" value to the 5 point Likert scale values; however we do not understand the difference between neutral and no opinion. If the survey is repeated with revision, we will include "no response" in the Mosaic version to make that version better approximate the E-Mail version.

The remaining comments target particular questions, and those question numbers are provided in parentheses:

- It is unclear what kind of tool may be of general interest. (Question 1.11)

  We agree, but expect such tools to emerge in due time. Therefore, we favor the present language.

- Need to define Levels 3 and 4 (science data product levels). (Questions 2.1 and 2.2)

  We agree and will either define those levels or delete the levels from the questions.

- Clarify if format is logical or physical. (Question 2.3)

  We agree and when the questionnaire is revised will clarify the format.

- What is "video annotation?" (Question 2.13)

  We agree and when the questionnaire is revised will explain video annotation.

- Say "wait ONLY 1 week" not "wait 1 week." (Question 2.18)

  We agree and when the questionnaire is revised will add the word "only."

- These statements are mutually exclusive as phrased, but the commenter would like to be allowed to agree with both. (Questions 2.23 and 2.24)
- The comment suggests a preference for automatic delivery (2.23) so no problem is apparent to this author. Also, the respondent is allowed to agree or strongly agree with one or both questions whether or not the phrasing of the questions suggests that one should prefer only one of the 2 options for delivery. Accordingly we do not see a need for revision.
- The comments identify questions 2.24 and 2.25 as unclear or ambiguous including a lack of understanding about "specifying what I want."
- We experienced difficulty in communicating these notions, and, not surprisingly, some respondents found the questions less than clear. Revision seems indicated.

We believe that whenever many respondents identified deficiencies in the same questions we should omit those questions from the reported results. Questions 2.1 and 2.2 would be the strongest candidates for deletion, but only 6 comments in 160 responses is judged as too few criticisms to justify their deletion. We retain all questions in the quantitative analyses because we believe that all are valid.

We generally conclude from the above comments that the instrument could benefit from some sharpening of its language. ESUS generated comments praising use of Mosaic and surprisingly few critical comments. We agree with the fundamental criticisms that were made, but stand by the present approach in light of the survey's goals and constraints. We believe that the instrument has generally been validated by respondents' comments and that ECS should use the results, provided that ESUS is properly labeled as a preliminary survey.

4.3 Quantitative Results and ECS-Relevant Comments

This section combines descriptive statistics about the responses to statements in Sections 1 and 2 of the questionnaire with a discussion of respondents' comments that considerably enrich and enliven the statistics.

Table 4-2 summarizes the responses to the background questions in Section 1. This table also summarizes respondents' requests for a report of the survey results, the frequency with which respondents contributed at least one comment, and the prevalence of response by E-Mail (rather than by Mosaic). We asked if respondents wanted a report on the results in order to compile a mailing list for this paper. 99 respondents (62 percent) requested a report, which we interpret as substantial interest in the issues, data, and interpretations contained in this paper.

Typical respondents are data product developers and consumers who research on regional and global geographic scales over time scales ranging from days to years, are receptive to the idea of adopting science tools that other scientists might develop, usually work with a few data sets at a time, and want a copy of this paper.

The mean responses to statements in questionnaire Sections 1 and 2 are appropriate to use whenever one needs the single best description of all respondents. However, examination of the frequencies in Table 4-2 or the histograms in Figures 4-1 through 4-9 immediately shows a wide spread in the responses to nearly every question. The presence of this spread suggests that anyone using the mean responses as the only descriptor of the respondents would be neglecting the extensive differences in opinion that will be surfaced in Section 4.4. Such spreads indicate that ECS should not refer to **THE** ECS science user community or **THE** ECS science user as if all science users were alike. These spreads attest to a diversity that needs to be better understood and incorporated into system design and management. Although the previous paragraph described the attitudes of the typical respondent, that characterization is over simplified and should be consumed cautiously.

Table 4-2. Summary for Questionnaire Section 1, etc.

Statements in Section 1, etc.	No	Yes	Mean Response
1.1 I am an instrument developer.	131	29	No
1.2 I am a data product developer.	61	99	Yes
1.3 I am a data product consumer.	19	141	Yes
1.4 The geographic scale of my primary research interest is local.	124	36	No
1.5 The geographic scale of my primary research interest is regional.	71	89	Yes
1.6 The geographic scale of my primary research interest is global.	32	128	Yes
1.7 The time scale of my primary research interest is decades.	84	76	No
1.8 The time scale of my primary research interest is years.	37	123	Yes
1.9 The time scale of my primary research interest is months.	53	107	Yes
1.10 The time scale of my primary research interest is days.	65	95	Yes
1.11 I plan to contribute a science analysis tool to ECS.	90	70	No
1.12 I would use science analysis tools that other scientists might contribute to ECS.	28	132	Yes
1.13 The number of data sets that I work with at one time is usually ONE.	150	10	No
1.14 The number of data sets that I work with at one time is usually A FEW.	57	103	Yes
1.15 The number of data sets that I work with at one time is usually MANY.	104	56	No
I provided comments.	105	55	No
I want a report.	61	99	Yes
I replied by E-Mail (not Mosaic).	47	113	Yes

The remainder of section 4.3 focuses on comments that shed light on opinions of the respondents regarding the desired ECS capabilities and design. These comments were stimulated by the statements in Section 2 of the questionnaire. A total 112 comments were received from 54 of the 160 survey respondents. In general, the comments are friendly, thoughtful, and provide useful insights based on the respondents' current work and past experience. The comments have been shared with the ECS design team and the ESDIS project in order to promote a better understanding of the ECS users' perspective.

In interpreting a user's comments, it is important to know whether this user's opinion is representative of many respondents or is a minority opinion. For this reason, this section also presents the overall distribution of responses to each of the survey items. This analysis is distinct from the cluster analysis presented in Section 4.4, which identifies and characterizes groups of like-minded individuals.

Rather than treating each of the 26 survey items (Section 2 of the survey) separately, we analyze related questions together. This yields 9 groups of questions.

4.3.1 Level 3/4 Grids

The first pair of questions concerns gridding of Level 3 and 4 data. Figure 4-1 shows the distribution of responses. A list of sample comments accompanies this figure, followed by an analysis of both.

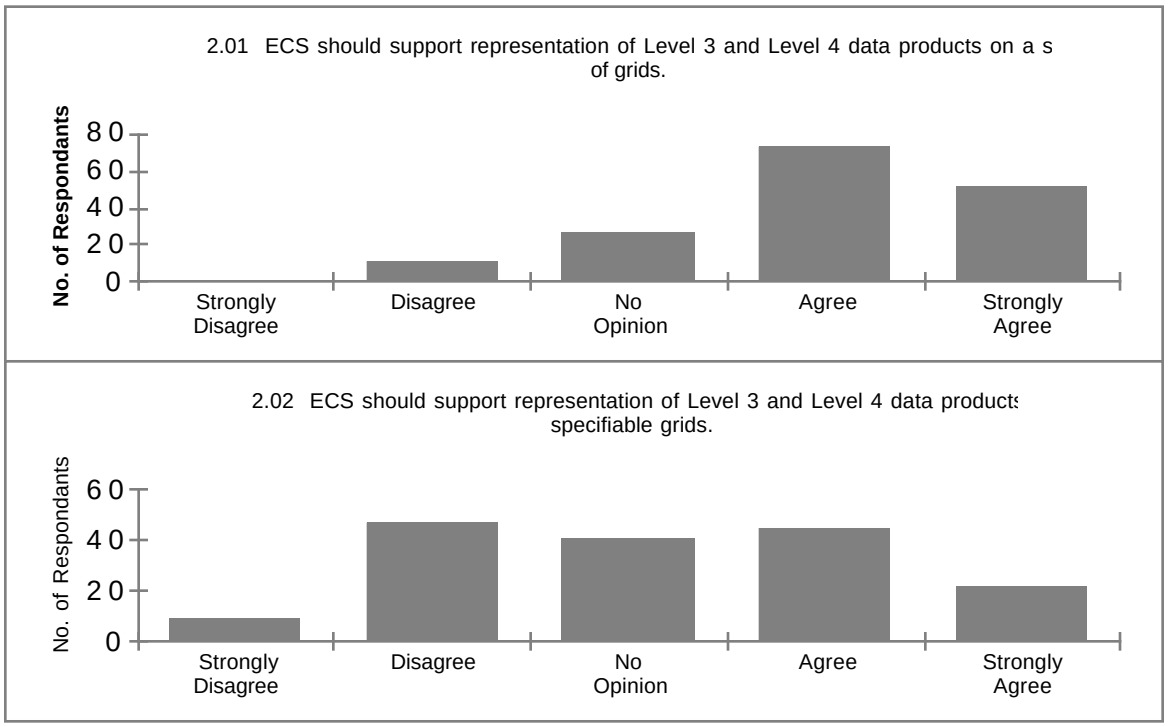

Figure 4-1. Frequencies for all Respondents (Statements 2.01-2.02)

Comments:

- I would be concerned if there was no user-specifiable grids. We have found that users want as much customization as is feasible.
- Provide conversion tools (that run transparently on the user's computer) to convert to the user-spec. grid.
- This is essential for comparing data from different products.
- ECS should support user-specifiable grids provided there is adequate information about interpolation methods etc. used.
- ECS should only deliver data on the grids used in its production. Changing grids/resolution is a science issue which ECS will not be qualified to do.

Analysis:

The responses in Figure 4-1 indicate strong support for a standard set of grids, with much less interest in user-specifiable grids. This might seem contrary to scientists' known preference for flexibility (as expressed in the first two comments, above). However, resampling of data has the potential for adding noise and artifacts. This is broadly recognized in the science community where many individuals have direct experience with the dangers of undocumented assumptions in regridding algorithms. This caution is reflected in the last two comments.

The third comment indicates an additional complication - while different data are best represented on different grids, comparisons require that the data be co-located. This is a scientific issue, rather than a data system issue, but its eventual resolution can have important implications for the data system's products.

4.3.2 Data Formats

Statements 2.3 through 2.6 concern the formats for distribution of ECS data to the users. Figure 4-2 shows the distribution of responses.

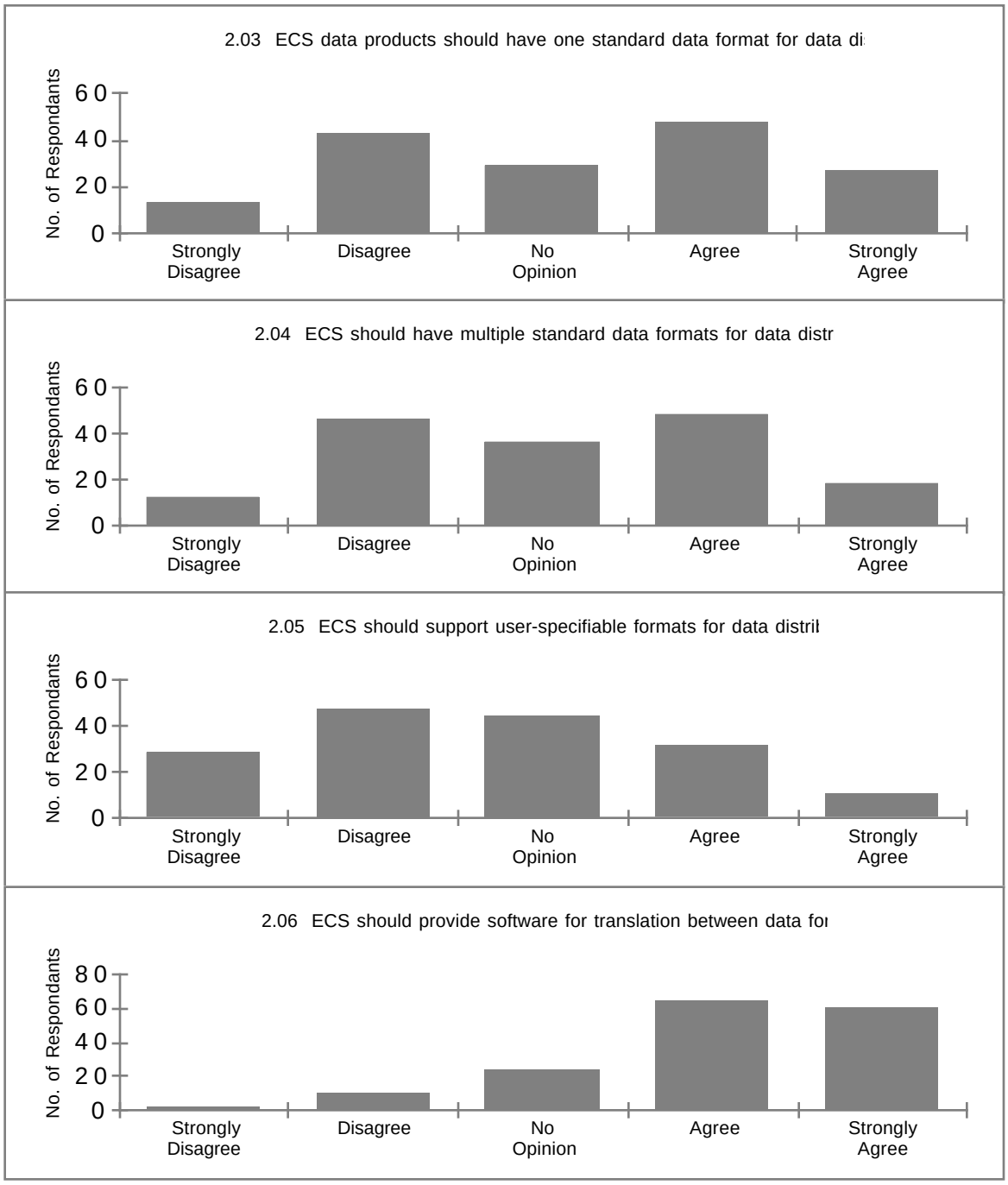

Figure 4-2. Frequencies for all Respondents (Statements 2.03-2.06)

Comments:

- Standard data formats have made life much easier for analyzing UARS data from various instruments, and I hope this is striven for in the EOS era, inasmuch as possible.
- It's impossible for scientists to agree on one (or a few) standard data formats. Aside from those that are just difficult, there will always be the scientist who comes up with something that cannot be dealt with by the standard format(s). Either a new data product or a new idea will be seriously hampered in its diffusion into the community by incredibly arcane gyrations to fit into the standard or will be lost entirely because it just doesn't fit.
- I do not support the idea of a single imposed data format for all data. The user should be allowed to select from a relatively small set of data formats. These should include ASCII and IEEE binary formats as well as something as awkward to use as HDF. I want to use the data, not employ an army of programmers to write programs that allow me to manipulate the data.

Analysis:

Except for statement 2.6, the responses in Figure 4-2 are ambivalent (and bimodal for items 2.3 and 2.4). This is reflected in the comments quoted above. The second and third comments express worries that rigid data formats will not accommodate the diverse types of data which must flow through the ECS and that these formats will make the data difficult to use. These are typical of most of the comments received.

The responses to statement 2.6 indicate that software tools are desired for translation between data formats.

4.3.3 Access to Scientists' Data

Statements 2.7 and 2.8 concern community access to data products developed outside the ECS. Figure 4-3 shows the distribution of responses.

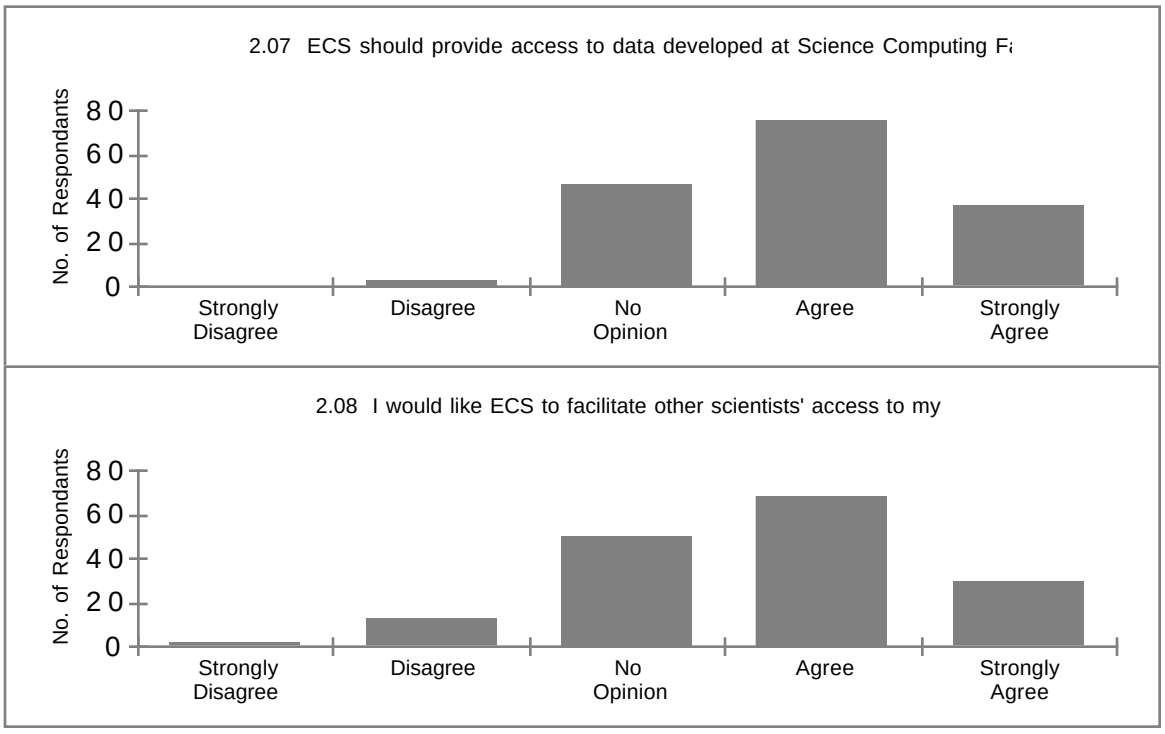

Figure 4-3. Frequencies for all Respondents (Statements 2.07-2.08)

Comments:

- The SCF's are fairly well funded and will represent a rather impressive compute and storage resource. Rely on them as a resource but, do not impose standards or schedules on them.

Analysis:

The responses in Figure 4-3 show strong agreement with the statement that access should be provided to data developed and/or residing outside of the ECS. The ECS extended-provider model provides for such access and was developed in response to user community inputs gathered early in the project. The present survey confirms these initial inputs.

4.3.4 Science Tools

Statements 2.9 and 2.10 concern software tools for data analysis and visualization. Figure 4-4 shows the distribution of responses.

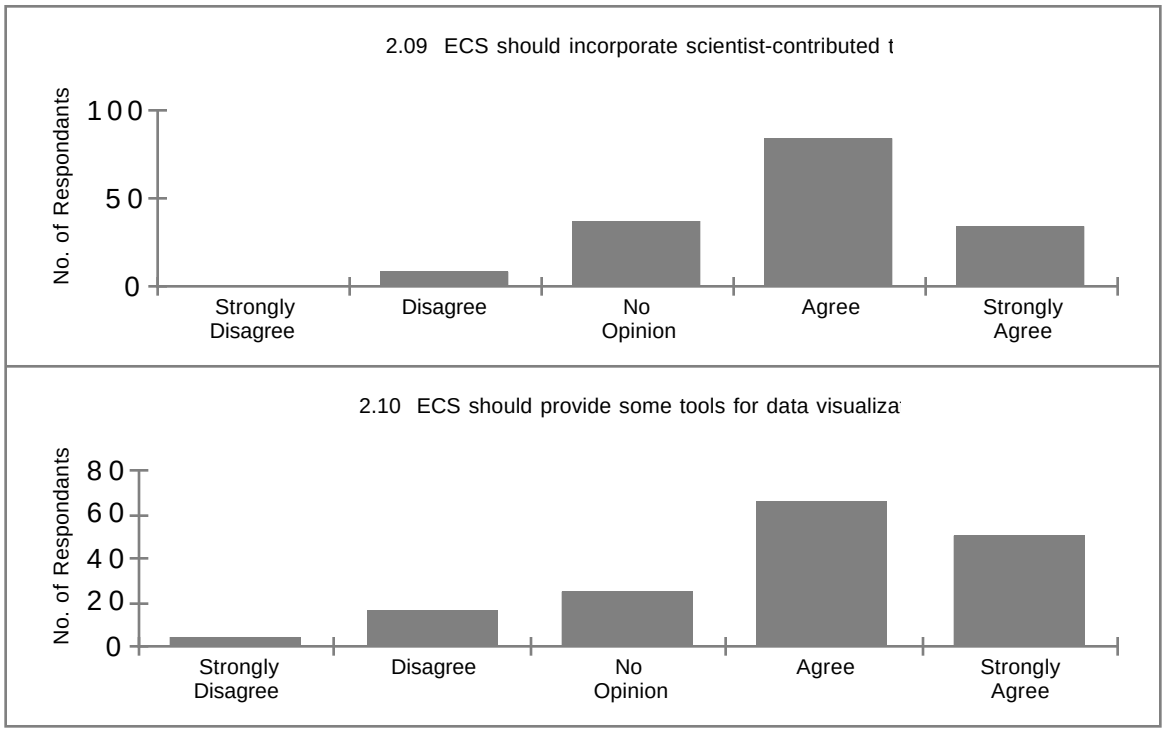

Figure 4-4. Frequencies for all Respondents (Statements 2.09-2.10)

Comments:

- Provide hooks to visualization packages that operate on the EOSDIS data formats such as NetCDF and HDF. Examples of such visualization packages are near infinity already.
- Many tools already exist. Easy access and summary of capabilities would be very useful.
- Interactive data examination is a big task, and good tools require a lot of work to develop. Also, there is considerable variation in what capabilities individual users need. Considering this, it would be very useful for ECS to provide non-proprietary tools for interactive data examination, especially visualization.
- [ECS should not provide data visualization tools] but should be compatible with commercially available tools.
- I am less inclined to want to add display software, especially with any degree of sophistication. My philosophy is to enable the access to the data so that the user can display it and further process it using their own system. However a good browse capability would be very useful.
- "Scientist contributed tools" works very well for STARLINK (UK astronomy) but has quality control implications.

Analysis:

These questions elicited many comments (only a subset of which are quoted on the previous page), suggesting that there are strong and well-formed opinions among the respondents. The distribution of survey responses in Figure 4-4 indicate a desire for a mechanism for scientists to share community software, as well as for some basic data visualization tool.

The comments indicate that the ECS should not attempt to compete with existing data visualization tools, since there is already a strong community and commercial investment in this area. The last comment suggests a valuable source for understanding requirements for sharing community-developed tools.

4.3.5 Multi-media Capabilities

Statements 2.11 through 2.13 concern multi-media (audio/video) capabilities. Figure 4-5 shows the distribution of responses.

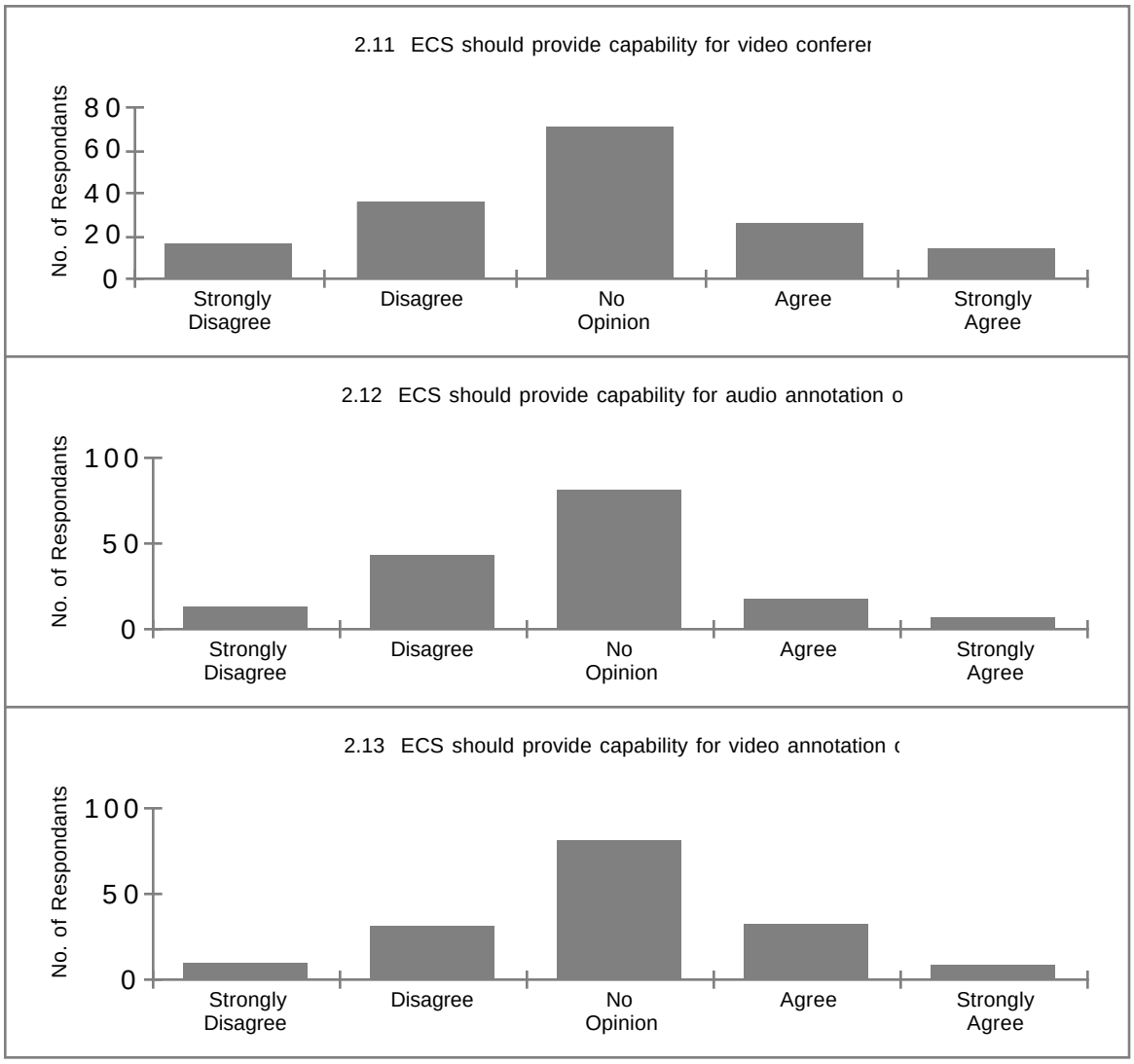

Figure 4-5. Frequencies for all Respondents (Statements 2.11-2.13)

Comments:

- Someone should certainly provide video conferencing. Does this mean that ECS should? Is that not pretty far from the key functionality of ECS?
- Within the MOSAIC framework [audio/video annotation of data] is an almost trivial pursuit.

- Audio or video annotation does not mandate the discipline that is required for written annotation; and the primary requirement for data in the ECS is that it be fully documented, validated, and understood!

Analysis:

Figure 4-5 shows that the responses to these statements were generally ambivalent, with the comments reflecting that multi-media capabilities are or would be developed outside of the ECS.

4.3.6 Data Distribution

Statements 2.14 and 2.15 concern expected media for data distribution. These statements did not specify the volumes of data being distributed. Figure 4-6 shows the distribution of responses.

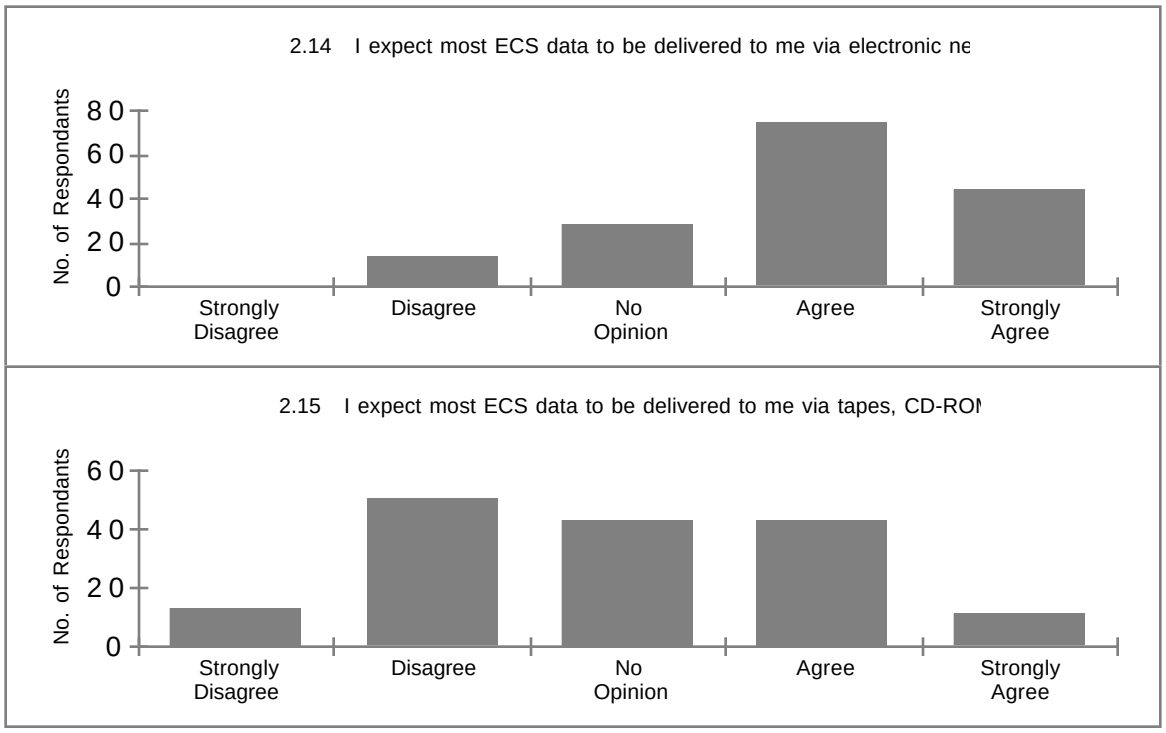

Figure 4-6. Frequencies for all Respondents (Statements 2.14-2.15)

Comments:

- For most purposes I would prefer electronic transfer, but for very big data sets I might actually prefer to receive a CD-ROM in the post, since it would avoid lots of work at my end in transferring it, keeping disc space open to receive it, avoid filling up my data link bandwidth, etc. Hence both are needed. The fact that in a few years disc storage and bandwidth will be much cheaper is not consolation - the data volumes will be much bigger, and anyway people will try more ambitious things, like getting the whole 10 year data set for an instrument whereas before they were content with a single year or a month, etc.
- One of the main issues that needs to be resolved in this context is subsetting of the data so that users can specify very specific data instead of having to digest multiple megabyte data files as is currently often the case.
- Preferred media will depend on size of sets. Electronic receipt of data will be difficult for large sets. For such sets, Exabyte tapes would be optimal for short term delivery, CD's for longer term.
- I would prefer to get data delivered or retrieved over the network because it is always faster but some data sets can be too large to comfortably send over the network and are easier to archive on tape or CD-ROM.
- As a non-US investigator, until trans-Atlantic networks improve it is vital that there are low band width equivalents to high band width services.
- If data are small volume, I expect it to be delivered to me via electronic networks, because in Japan, networks do not have large capacity.
- Unless the networks get upgraded (especially in Canada to T1 and T3) it would seem that some other medium is preferable for GBytes of data that will be required to analyses.

Analysis:

Figure 4-6 shows that most respondents agreed that network distribution of data would be preferred. The majority of the comments indicate a realization that distribution of large data sets over networks is still impractical. A number of commentors observed that the ability to precisely specify the desired data is critical to cost effective (and timely) data distribution. Finally, the last three comments reflect a concern by non-U.S. investigators that their links will be unable to accommodate high-rate data transfers.

4.3.7 Response Times for Data Delivery

Statements 2.16 through 2.18 concern tolerable times to wait for data delivery. Figure 4-7 shows the distribution of responses.

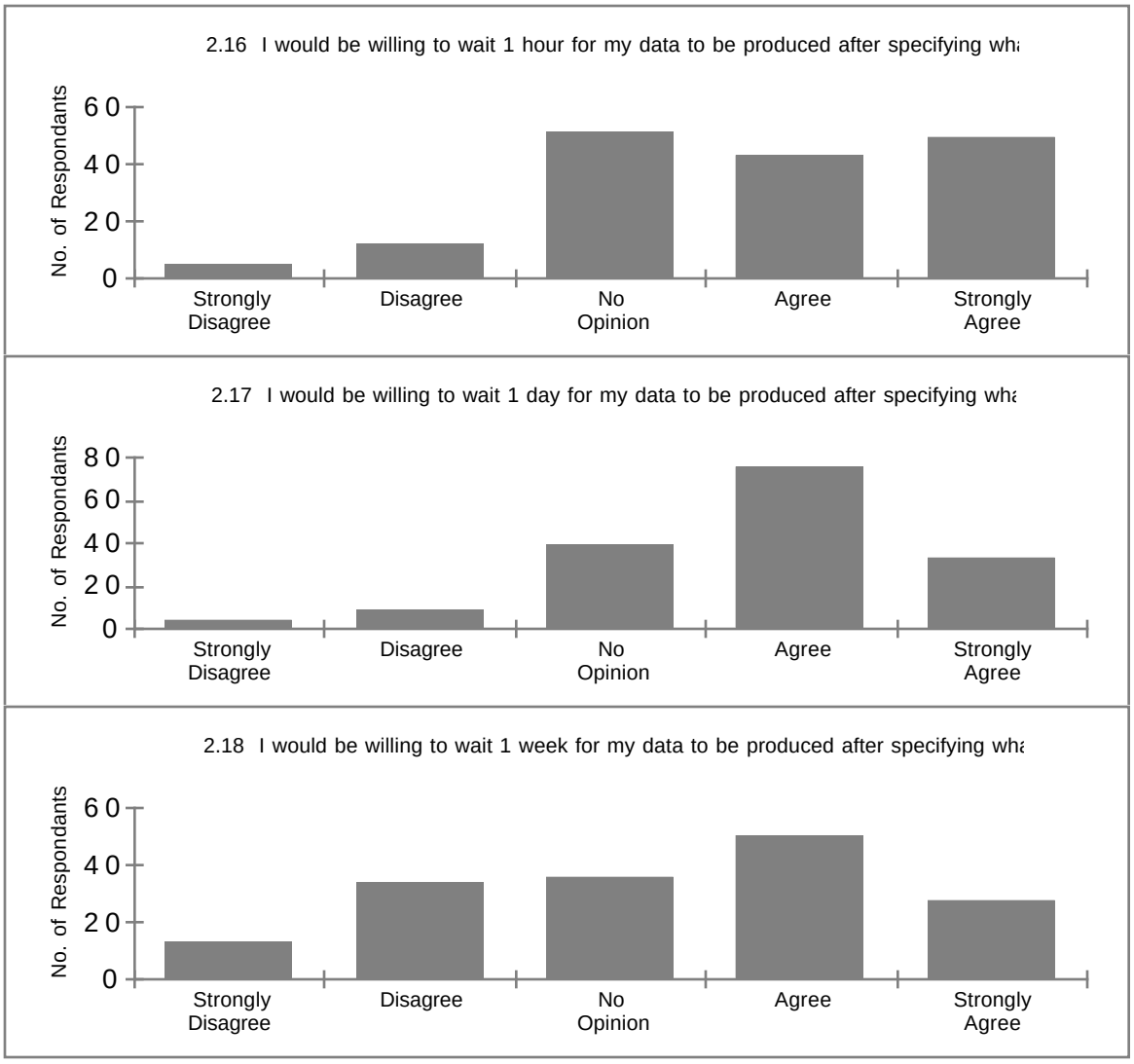

Figure 4-7. Frequencies for all Respondents (Statements 2.16-2.18)

Comments:

- The answer to [these] questions is dependent on data volume:
  a. For small subsets (< 1 MB), where I have specified at least a selection from a single granule, I would be willing to wait 5 minutes to 1 hour.
  b. For samplings from a few granules, or a few granules, I would be willing to wait 1 hour.
  c. For specification covering many (10 or more) granules, I would be willing to wait 1 day to 1 week.

- For less than 1 Mb of data, I want instant access. For 1 to 50 Mb, I could wait an hour. I would wait 1 day for 50-500 Mb. I would wait 1 week for data sets over 500 Mb.
- What turn a round I expect depends on volume. If browsing or checking out minor points I want quick access; with larger amounts I don't mind waiting a day, or could get used to putting in an order before going home at night, knowing that it will have arrived by next day. With huge orders I would be content with a week or more and might prefer them on CD.
- As long as retrieving information about the data is immediate. In other words it is ok to wait for the data to arrive once the user has specified exactly what parts of which data sets he or she wants.
- I don't mind waiting overnight for a big data set. I tend to lose my train of thought if it takes a long time to get to a data set. And if what I get doesn't turn out to be what I wanted, another week is a long time.
- [I would be willing to wait 1 day] for most of my data, but when a volcano is erupting, I'd like the data as fast as possible.
- Depends upon the nature of the request. If it is requested as part of a decision-making process about more data, then even 1-hour may be too long. If it is a request for a large volume of data that will be subject to intensive data analysis, then 1-week is OK.

Analysis:

The responses in Figure 4-7 indicate that the acceptable data delivery time is somewhere between 1 day and 1 week for most respondents. However, the comments clearly indicate that a uniform response time covering all cases is undesirable. As the first three comments indicate, acceptable delivery time depends on the volume of data to be delivered. This is consistent with the expectation that large deliveries should be on media, rather than through networks.

The next two comments indicate that longer delivery times are more acceptable if the desired data can be specified with a high degree of confidence and the system will deliver those data which were specified on the first try (i.e., the data distribution system must reliably deliver the specified data). Many users of current systems have experience with ordering data and waiting a week for delivery, only to find that the tapes are unreadable, in a strange format (and not accompanied by software for reading the format), or are missing some of the expected data. In such cases, 1 week delivery times translate into much longer times to acquire the required data.

The last two comments point to specific instances in which scientists will want quicker data delivery. Because the system will be unable to anticipate the intended application of the data, scientists must be able to specify the urgency (or maximum acceptable delivery time) for a given data delivery.

4.3.8 Long-term Consistency vs. Latest Algorithms

Statements 2.19 through 2.22 concern trade-offs between data with long-term consistency of processing vs. use of state-of-the-art algorithms. Figure 4-8 shows the distribution of responses.

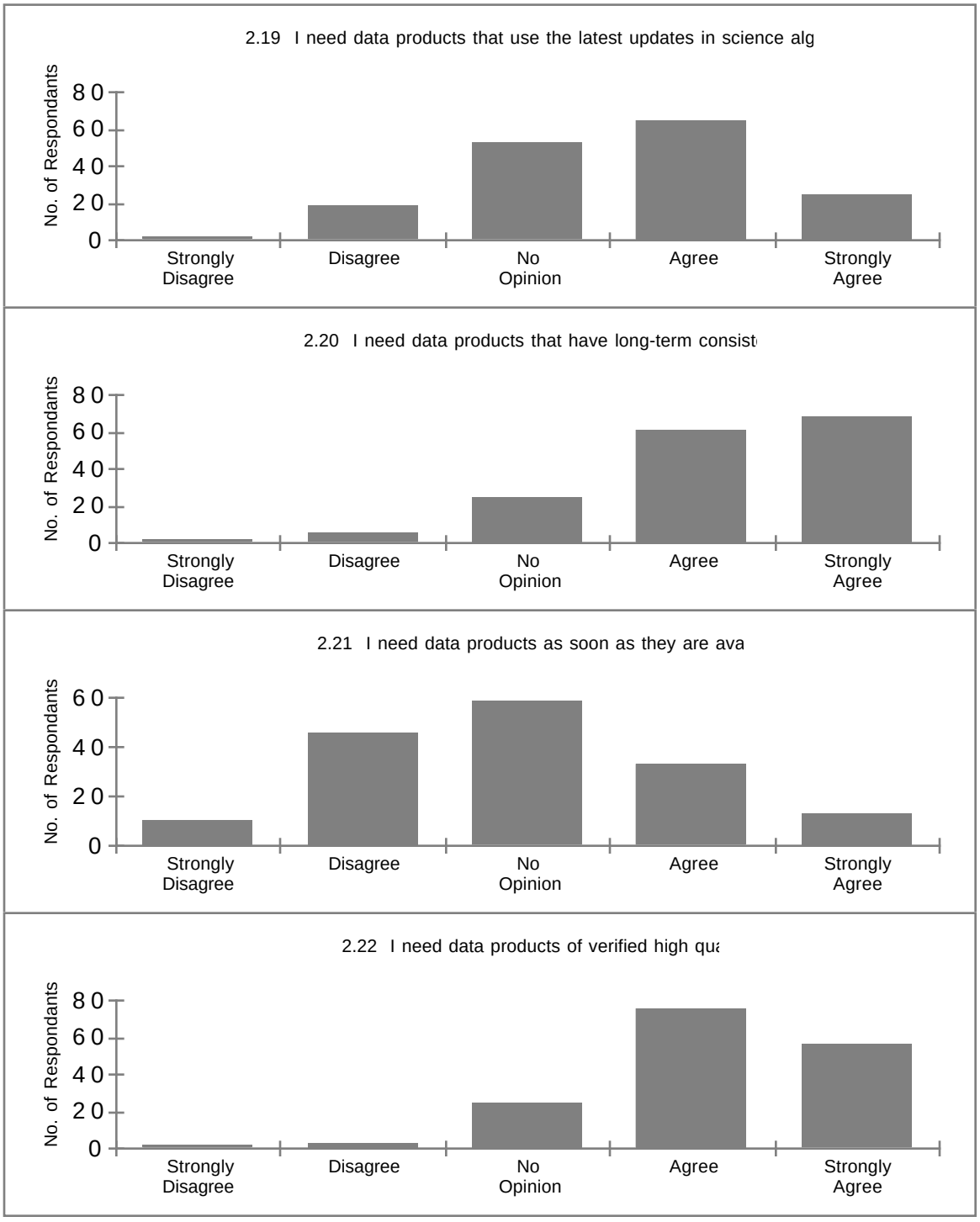

Figure 4-8. Frequencies for all Respondents (Statements 2.19-2.22)

Comments:

- Data once produced and standardized should be supported for long periods especially in the global change community. It is in some instances [it is] more advantageous not to have the latest algorithms being applied because one might only be interested in relative changes. Many instances have been reported where it became clear after a long study that the observed changes in certain years were simply due to changes in the methods by which the observations were processed.
- The experience with past satellite instruments is that it takes many iterations to converge on a retrieval algorithm that produces reliable, high quality data. While rapid dissemination of early versions may produce some useful feedback, in my opinion the disadvantages overwhelm this effect: the effort wasted and misconceptions propagated do a lot of damage.
- I need highly documented quality control reports so that I can decide for myself whether the quality matches my expectations or is suitable for what I intend to use it for.
- It isn't contradictory to want some of the data processed with the latest algorithm, but also all the data processed with the same algorithm - even if it has known (or even unknown) limitations.
- I expect to use data for a number of purposes. Some will require near-real time data; others will require long-term consistent data sets e.g., studies of interannual variability.
- I expect to have a range of data uses. While much of the time I will be willing to wait for higher quality data, I can anticipate occasionally wanting some data products in near real time, such as when conducting field measurements.

Analysis:

Figure 4-8 shows strong agreement about the need for data products of verified high quality and long-term consistency. This is also reflected in the first three comments. There is also considerable desire for data products that were processed using the latest algorithms. In principle, it is possible to have both long-term consistency and latest algorithms, but this is often not practical for long-term data sets requiring extensive processing.

The need for access to data products as soon as they are available does not appear to be as strong, though the comments listed above (and many others, not included above) show that the respondents feel that there are instances when they will need data access in near-real-time.

4.3.9 Data Access

Statements 2.23 through 2.26 concern methods for user access to ECS data. Figure 4-9 shows the distribution of responses.

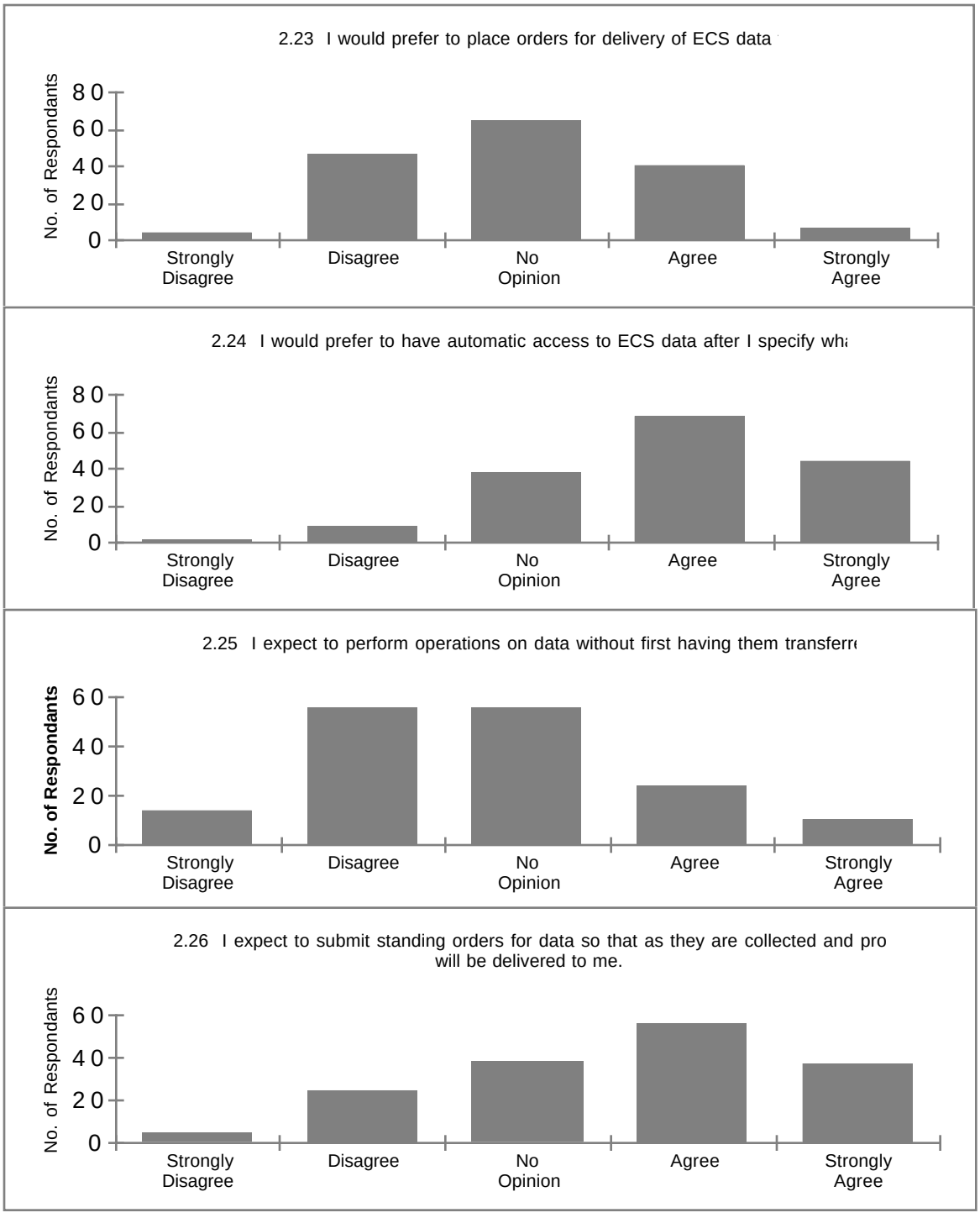

Figure 4-9. Frequencies for all Respondents (Statements 2.23-2.26)

Comments:

- It is important that users have electronic access to data. Often one wants to experiment with data prior to ordering large volumes of it. Long delays in acquiring the data will stand in the way of such experimentation.
- There is a need for tools to subset data to avoid the need to transfer unwanted data . . . This way responsible users could greatly reduce the volumes transferred, and do the job in a flexible way that is adaptable to all kinds of selection criteria that would not be dreamt of in advance.
- Certain sorting/ filtering/ extraction operations should be performed at the ECS end, within limits of I/O and CPU load. This should lower the burden on the network and the recipient machine's I/O channels.
- It is essential that flexible data sub-setting tools are provided by ECS so that data transfer volumes are kept reasonable and users are not overwhelmed with data which they do not require.
- Part of our research plans would require some change detections run on data and automatic notification to us.

Analysis:

The responses in Figure 4-9 indicate a preference for automatic access to data, rather than the more traditional mode of ordering data for later delivery (statements 2.23 and 2.24).

The responses to statement 2.25 indicate that most users do not plan to perform operations on the data before transferring them to their facilities. This is in disagreement with the comments which indicate that users would like to be able to subset or subsample the data before transfer. These (and other similar comments) recognize the advantages to both the system and the user in reducing the volumes of data to be transferred. Either these comments represent a minority opinion (though there were no comments to the contrary), or many users responding to statement 2.25 do not consider subsetting as "performing operations on data" (i.e., they may consider subsetting to be operations performed by the system).

Most users expect to utilize standing orders in order to access at least some of their data.

4.4 Cluster Analysis to Identify Like-Minded Scientists

4.4.1 The Method and the Resulting Clusters

Cluster analysis was performed in order to identify groups of scientists who responded similarly to the questions in Section 2 and therefore have similar visions for ECS and their use of ECS. Ideally the clusters would be very tight so that most of the diversity among the respondents would be preserved when aggregating the 160 respondents into a smaller number of groups. Each cluster should differ considerably from each other cluster, and the data points (respondents) within each cluster should be highly similar (close) to all other points in that cluster. Cluster analysis treats each respondent as a point in a 26 dimensional space with one dimension for each question. When asked to determine some number of clusters, such as 10, the technique iteratively searches for ways to group the 160 individuals into 10 clusters while minimizing the total Euclidean distance

among the members of each cluster. By minimizing those distances the analysis forms groups that are as compact as possible in 26 dimensional space. This analysis was performed using ANTHROPAC 5.0 software.

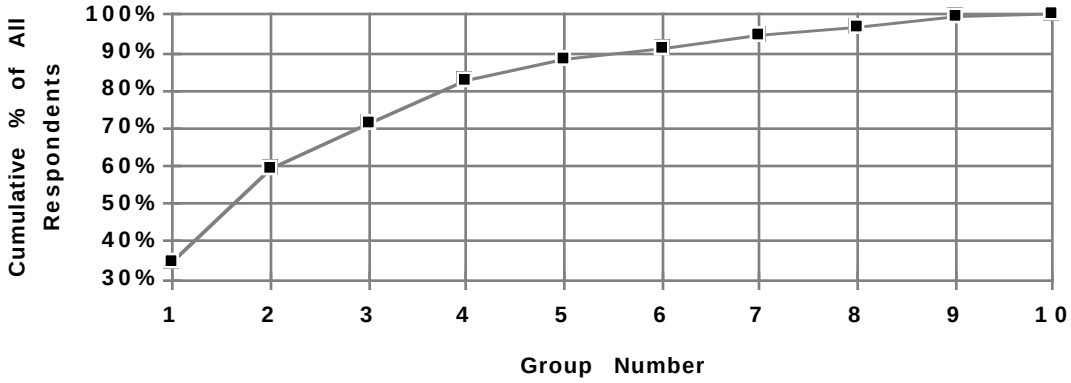

Figure 4-10. Cumulative Percent of Respondents in Clusters (Groups)

Figure 4-10 shows that use of cluster analysis to identify 10 clusters (groups) has produced more groups than necessary. 88 percent of all respondents are members of the five largest group (1, 2, 3, 4, and 5 in order of decreasing size). This suggests that ECS could validly focus its efforts upon establishing dialog with these largest groups and Section 4.4.2 also concentrates on characterizing the largest groups. Groups are discussed in order of decreasing size.

4.4.2 Profiles of Identified Groups of Scientists

In this section we capture the diversity among the scientist respondents by describing the characteristics of each cluster (Group 1, Group 2, etc.). Each of the Figures 4-11 through 4-20 shows one group's ranges of response to the questions in Section 2. The solid diamonds within the ranges mark the each group's mean response. The means effectively summarize the response profile that is typical of each group. Figure 5-21 shows the ranges of response for all 160 respondents. Table 4-3 shows that all groups have from 1 to 55 members.

Table 4-3. Sizes of Groups Identified by Cluster Analysis

Group	1	2	3	4	5	6	7	8	9	10	All
Number of Members	55	40	19	18	9	5	5	4	4	1	160

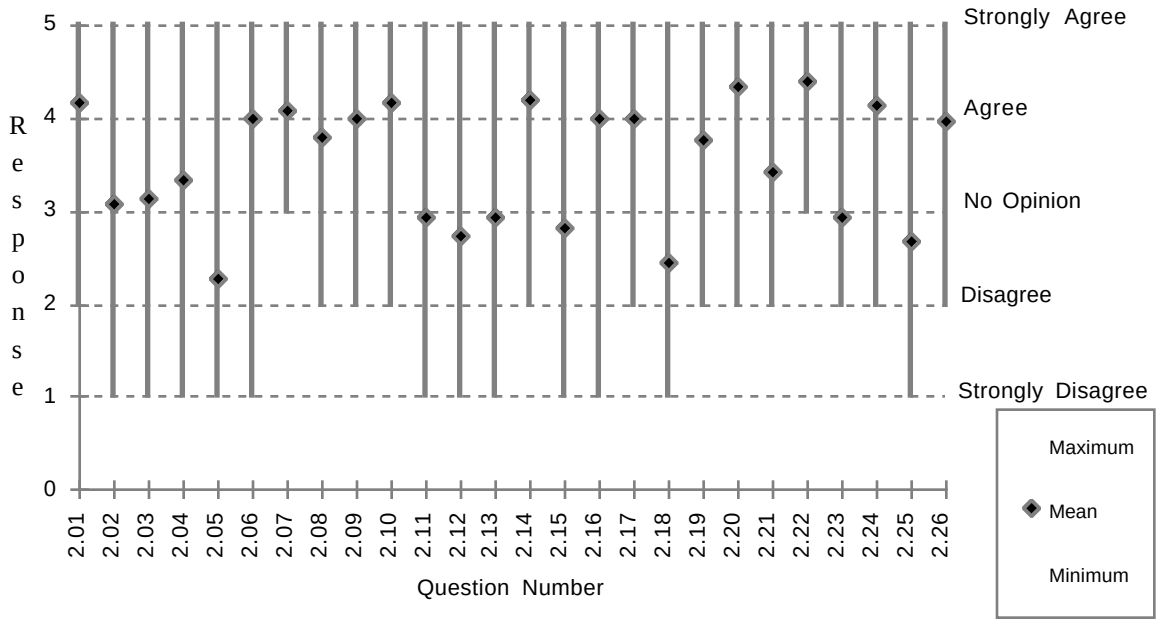

Figure 4-11. Group 1's Ranges of Response for Section 2

With 55 members, Group 1 is the largest group and comprises more than one-third of all respondents. Compared to all respondents this group is under represented by

- Data product developers
- Scientists focusing on decade time scales
- Scientists who work with 1 or many data sets at a time

The members of Group 1 have the following preferences:

- ECS should not support user-specifiable data formats
- ECS should
  - Support a standard set of grids and translation between data formats
  - Provide access to SCF data
  - Facilitate other scientists' access to their data
  - Incorporate scientist-contributed tools
  - Provide some tools for data visualization

Group 1 members

- Expect to receive data delivery via electronic networks
- Are willing to wait a day or less for data to be produced, but would not wait a week
- Need data products that
  - Use the latest updates
  - Are of verified high quality
  - Have long-term consistency
- Prefer to have automatic access to ECS data
- Expect to submit standing orders for data

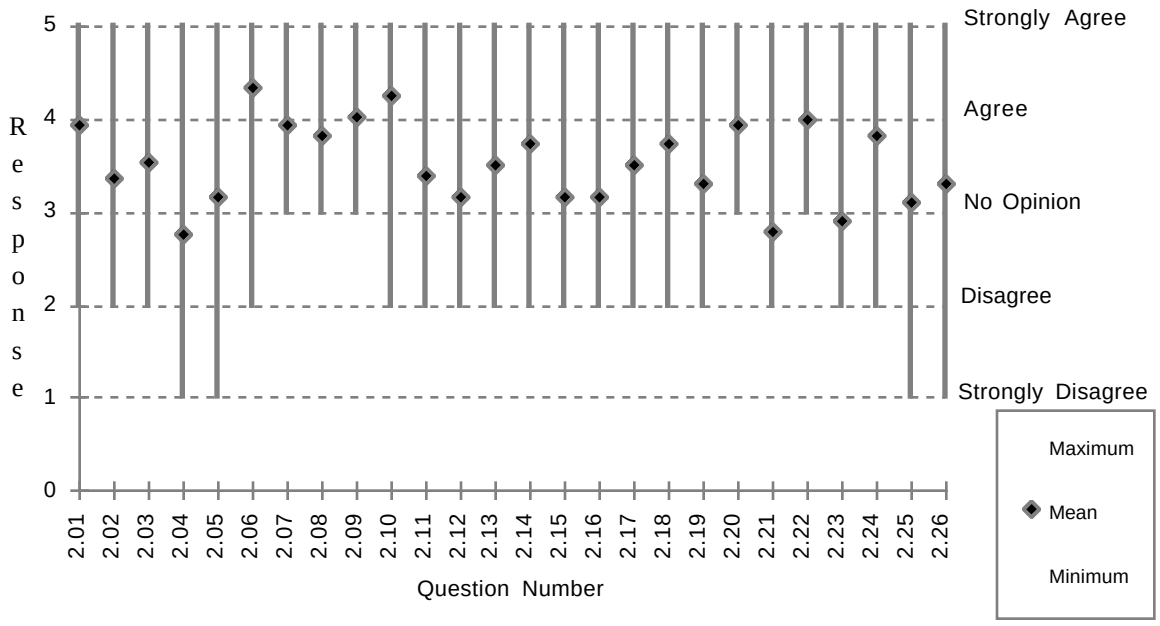

Figure 4-12. Group 2's Ranges of Response for Section 2

Group 2 accounts for 25 percent (40) of the respondents. Compared to all respondents, Group 2 is under represented by:

- Instrument Developers

Group 2 is over represented by:

- Data product developers
- Scientists who usually work with 1 or many data sets at a time

The members of Group 2 believe that ECS should

- Support a standard set of grids and translation between data formats
- Distribute one standard data format
- Provide access to SCF data
- Facilitate other scientists' access to their data
- Incorporate scientist-contributed tools and
- Provide some tools for data visualization.

Group 2 members

- Expect data delivery via electronic networks
- Are willing to wait 1 week for data to be produced
- Need data products with
  - Long-term consistency and
  - Verified high quality
- Prefer to have automatic access to ECS data after they specify what they want.

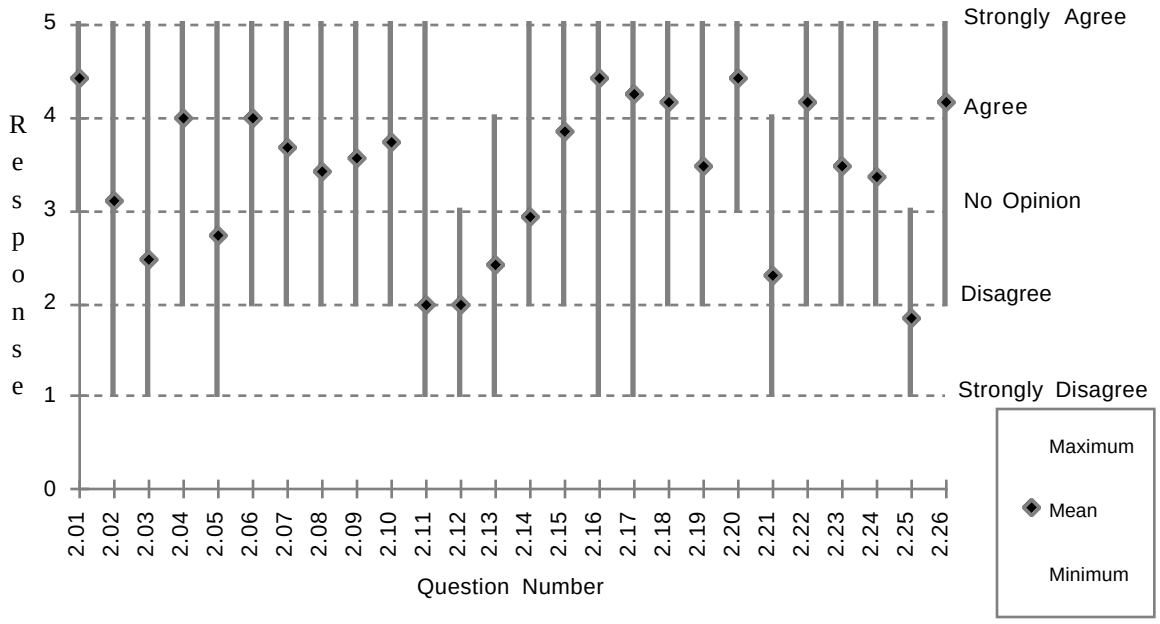

Figure 4-13. Group 3's Ranges of Response for Section 2

Group 3 accounts for 12 percent (19) of the respondents. Group 3 is under represented by members who

- Usually work with many data sets at a time

Group 3 is over represented by members who

- Are instrument developers
- Focus on local or regional geographic scales
- Usually work with 1 data set at a time

The members of Group 3 believe that ECS should

- Support a standard set of grids and
- Distribute data with multiple standard formats
- Support translation between data formats
- Provide access to SCF data
- Incorporate scientist-contributed tools
- Provide some tools for data visualization
- Should not distribute data with one standard data format
- Should not provide capability for video conferencing or audio or video annotation of data

The Group 3 members

- Expect data delivery via hard media and
- Are willing to wait from 1 hour to 1 week for data to be produced
- Need data products with
  - Long-term consistency and
  - Verified high quality
- Do not need data products as soon as they are available

Group 3

- Expects to submit standing orders for data
- Does not expect to perform operations on data without transfer

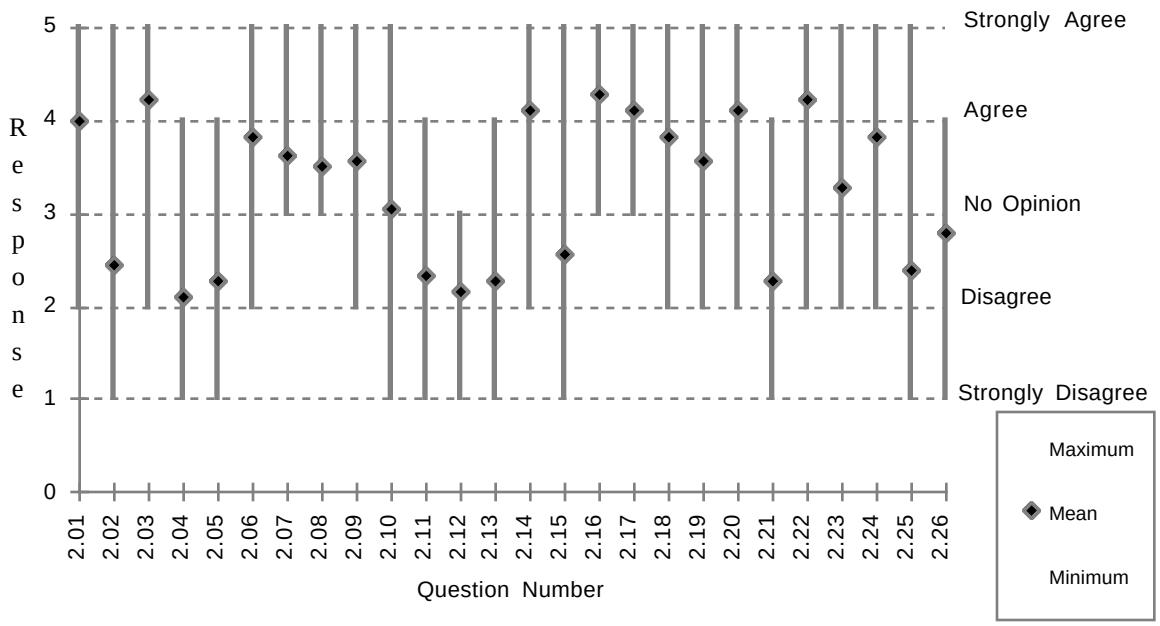

Figure 4-14. Group 4's Ranges of Response for Section 2

Group 4 accounts for 12 percent (18) of the respondents. Members of Group 4 are under represented by

- Instrument developers
- Local and regional researchers
- Scientists who plan to contribute a science analysis tool to ECS

Group 4 is over represented by

- Scientists who usually work with a few data sets at a time.

The members of Group 4 believe that ECS should

- Support of a standard set of grids
- Distribute data with one standard format
- Support translation between data formats
- Provide access to SCF data
- Incorporate scientist-contributed tools

The members of Group 4

- Prefer to have automatic access to ECS data
- Do not expect to perform operations on data without transfer
- Oppose distributing data with
  - Multiple standard formats
  - User-specifiable data formats
- Believe that ECS should not provide capability for
  - Video conferencing
  - Audio and video annotation of data
  - User-specifiable grids
- Expect data delivery via electronic networks
- Are willing to wait from 1 hour to 1 week for data to be produced
- Need data products that
  - Use the latest algorithm updates
  - Have long-term consistency
  - Are of verified high quality
- Do not need data products as soon as they are available

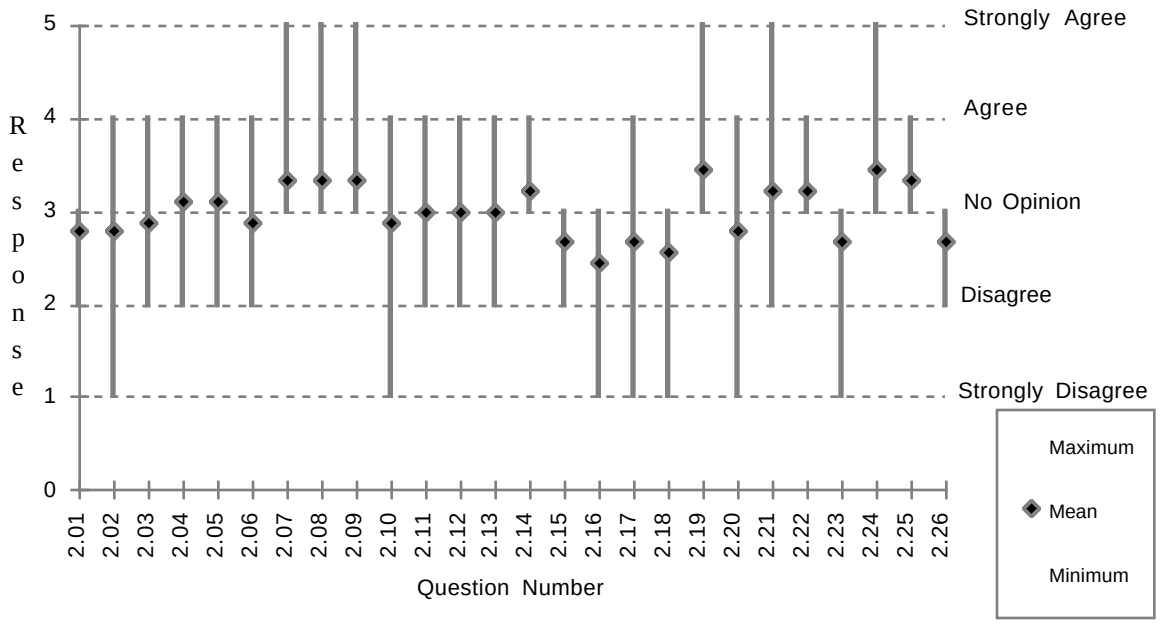

Figure 4-15. Group 5's Ranges of Response for Section 2

Group 5 accounts for 5 percent (9) of the respondents. Compared to all respondents, the members of Group 5 are under represented by

- Local researchers
- Focus on research over decades
- Work with 1 or a few data sets at a time

Group 5 is over represented by

- Work with many data sets at a time

The average member of Group 5 either did not respond or had no opinion on all questions in Section 2 except for disagreeing with the statement that "I would be willing to wait 1 hour for my data to be produced . . . "

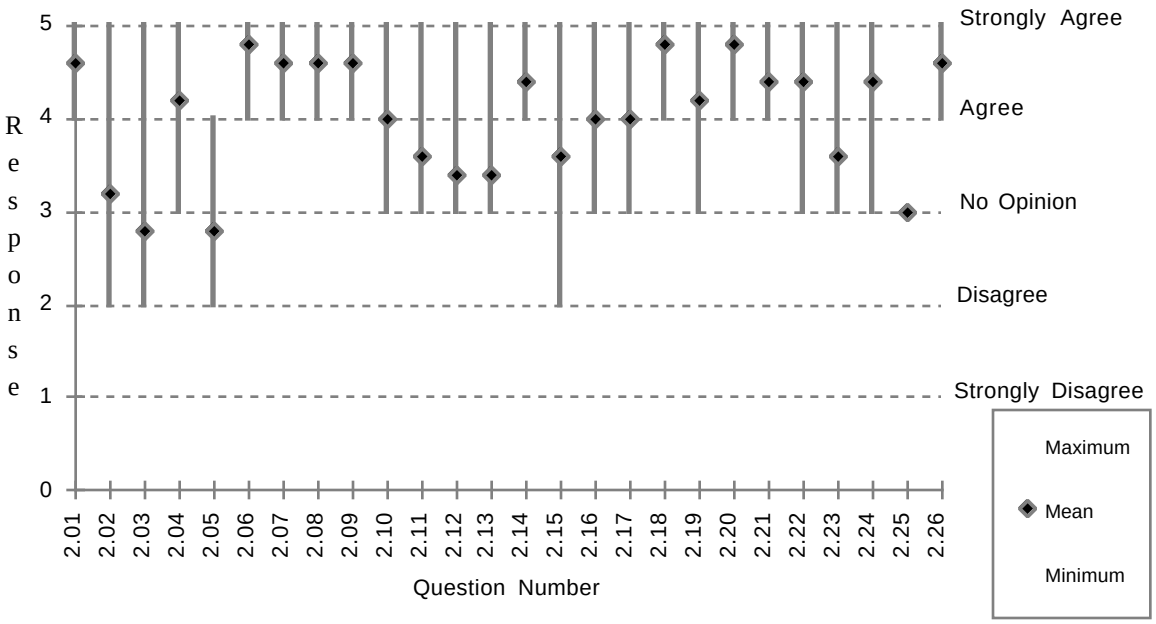

Figure 4-16. Group 6's Ranges of Response for Section 2

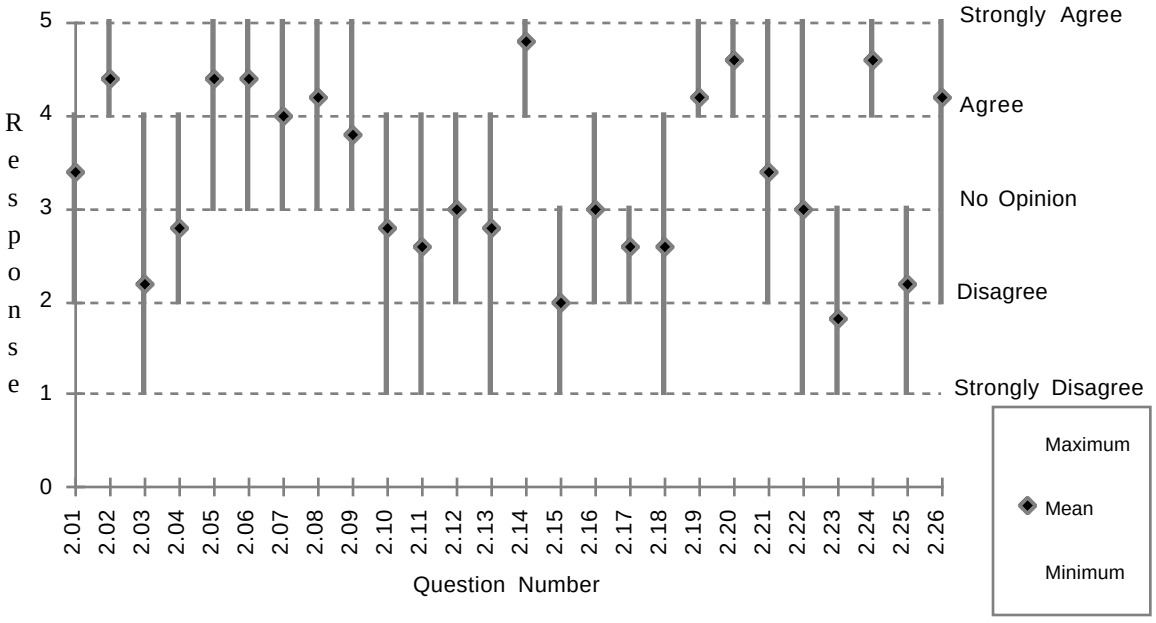

Figure 4-17. Group 7's Ranges of Response for Section 2

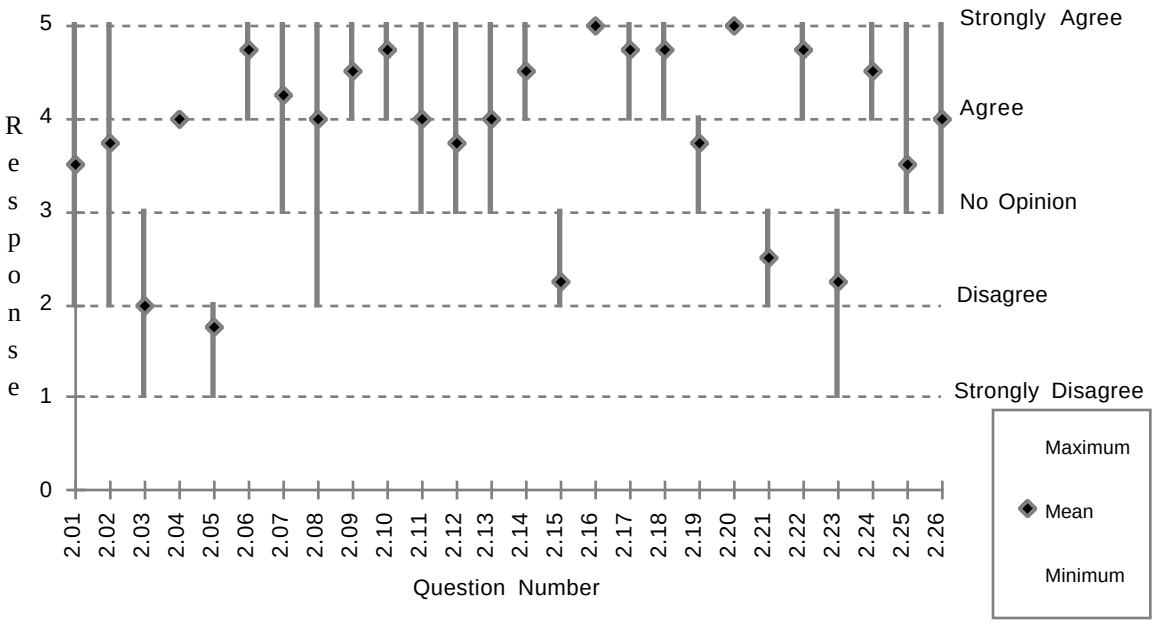

Figure 4-18. Group 8's Ranges of Response for Section 2

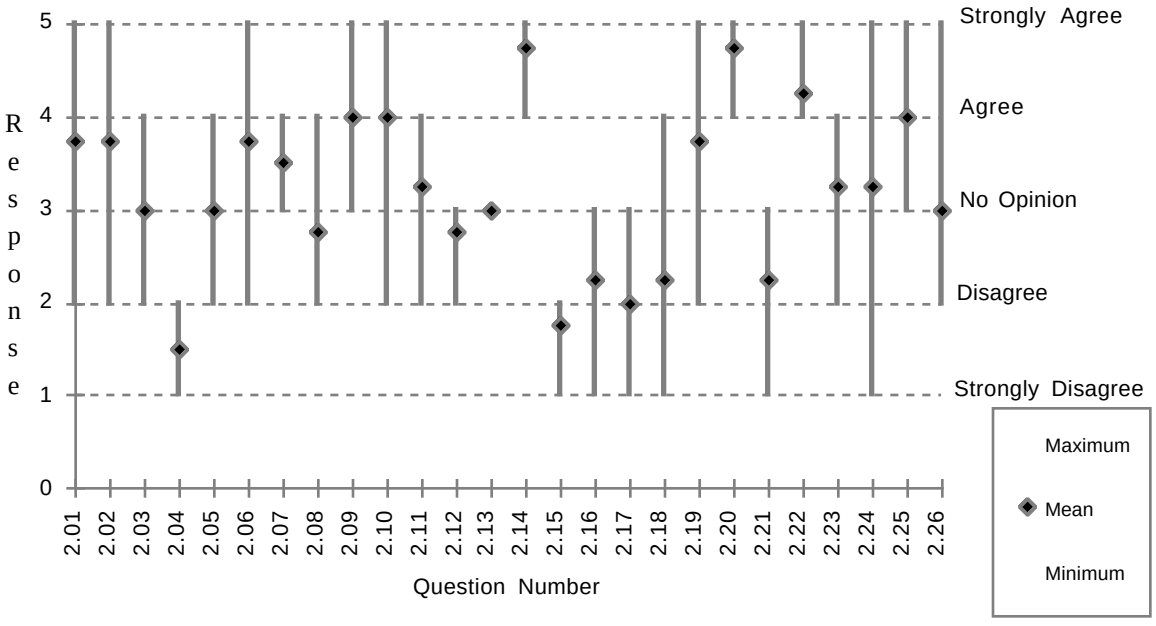

Figure 4-19. Group 9's Ranges of Response for Section 2

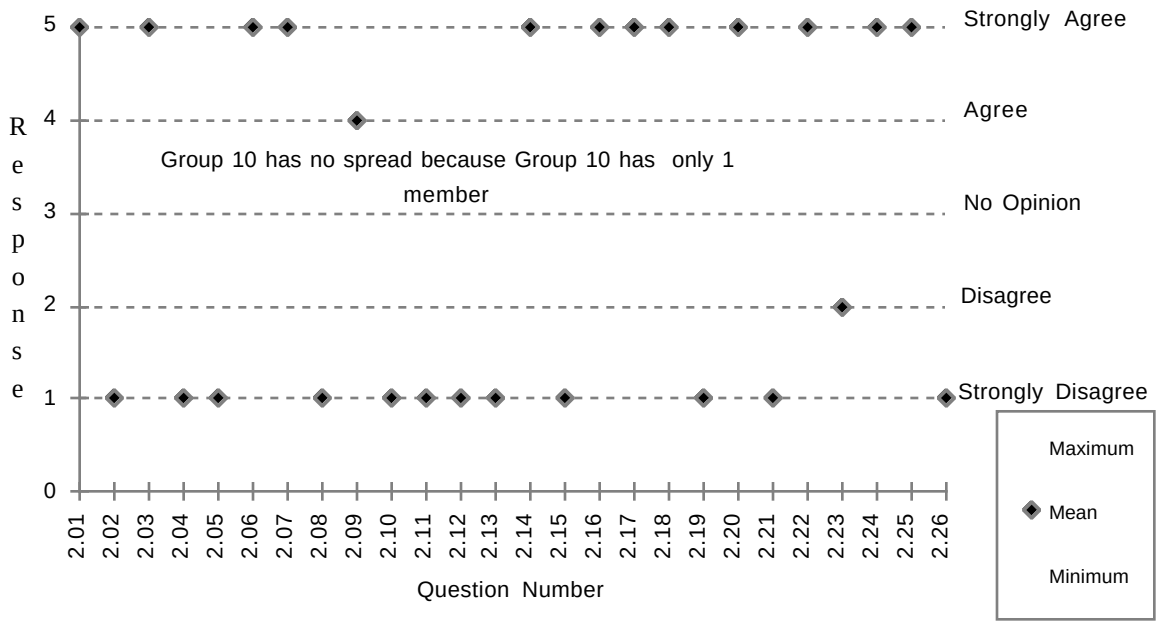

Figure 4-20. Group 10's Ranges of Response for Section 2

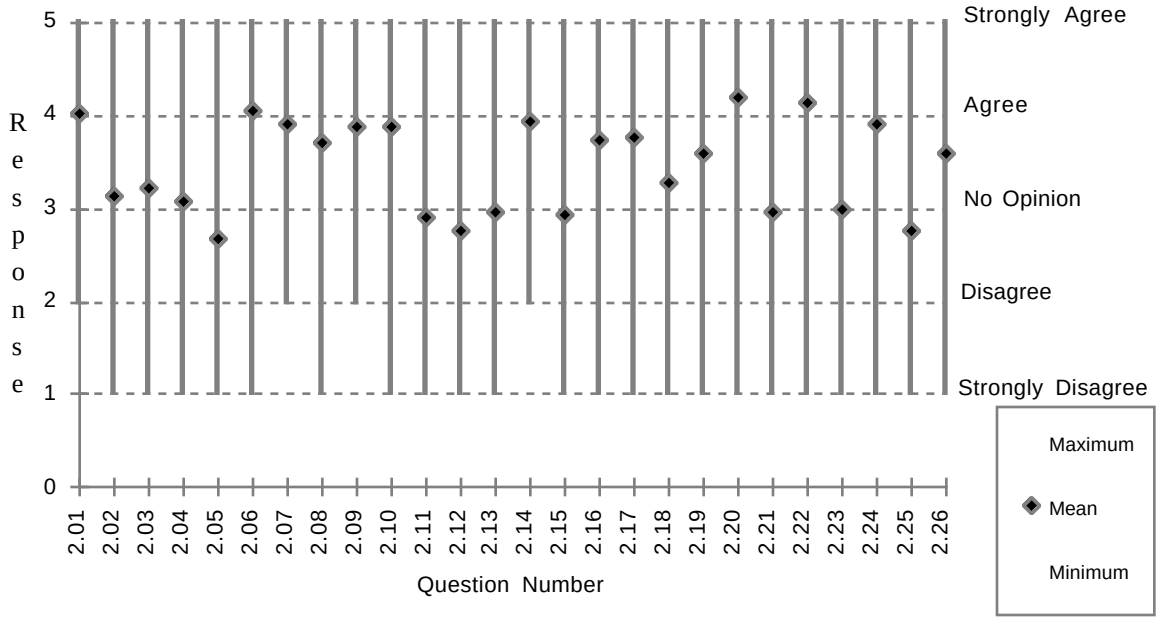

Figure 4-21. Ranges of Response for All Respondents

Figure 4-21 shows that the responses for all respondents have great spread. Except for three questions the range in responses span the entire 5 point Likert Scale.

4.4.3 User Diversity with Subcultures

This analysis examined the possibility that the respondents are members of one or more subcultures. Subcultures are variants of a culture where culture is defined as "learned categories of experience conventionally associated with learned plans for action"[1]. When restated for Earth system scientists the issue becomes whether or not groups of scientists share ideas about ECS and their use of ECS and, therefore, are members of a subculture. If a group of respondents are members of a subculture, then there is a culturally-correct answer to each question from the perspective of that subculture. Also, completely-knowledgeable members of a subculture would know the correct answers to all questions and respond with those correct answers. Less-knowledgeable members would be forced to respond randomly to questions about which they know little. Cultural consensus analysis estimates each respondent's knowledge of a subculture and estimates the subculturally-correct answer to each question (answer key). The most knowledgeable member of a subculture would be the best candidate for the representative of that subculture (translated as a group of like-minded scientists). Different subcultures accept differing truths in that each recognizes a different set of "correct" answers to the questions.

Cultural consensus analysis was performed using ANTHROPAC 5.0 software. The analysis of the 160 respondents showed that the responses violate the assumptions and that the respondents do not share a subculture. That failure is to be expected given the user diversity that this study was designed to explore.

We repeated separate cultural consensus analyses for each of the more homogeneous Groups 1, 2, 3, 4, and 5 that emerged from the cluster analysis in Section 4.4.1. The responses for each group showed cultural consensus, which supports the validity of the cluster analysis. The members of each of the groups appear to share a subculture relative to the questions in Section 2 of the questionnaire. However, Group 2 only marginally passed the criterion that is required for the conclusion that the members have one subculture. By identifying subcultures ECS can simplify the problem of responsiveness to many science users to the lesser problem of responsiveness to a few science-user subcultures.

The analyses for each of the groups estimated the knowledge levels of all members. This showed that at least one member except, for Group 2, possesses a high level of knowledge (0.8 or greater on the 0.0-1.0 knowledge scale) about that group's subculture. Group 2 had one member with a 0.72 knowledge level, which qualifies as competent but not particularly expert. Those experts will be prime candidates as group spokespersons or representatives of each groups. Although these groups are more homogeneous than the total population of respondents, each group also has members with low knowledge levels (0.5 or less). This observation underscores the desirability for ECS to work most closely with knowledgeable group members.

---

[1]Modern Cultural Anthropology, Second Edition, Philip K. Bock, N.Y:Alfred A. Knopf. 1974.page 448.

The cultural consensus analysis also estimated the culturally correct answers to each question. (answer key) Those answer keys approximated the group means for the clusters reported in the previous section. Although further work is needed before we can make a firm recommendation, we suspect that the answer keys determined by cultural consensus analysis will better characterize each group than the averages of the members' responses. We place less confidence in the means than the answer key because the means are calculated by equally weighting highly and lowly knowledgeable members. The estimated answer keys weight the responses of each respondent by his or her inferred knowledge level. As an example of a divergence between the means reported in the previous section, we concluded previously that Group 5 was undecided on all but one question. The cultural consensus analysis concludes that Group 5 is undecided on all questions.

5. Potential Applications of the Results to ECS Development

5.1 Generally Increase ECS Awareness of Diversity among Science Users

The ranges of response to questionnaire Section 2 show wide diversity among ECS science users. At least one person chose every possible answer to nearly all questions. But that great spread in responses does not imply a lack of structure in the response data (and ECS science user characteristics). Section 4.4.1 demonstrated that 88 percent of the 160 respondents can be captured by postulating that each respondent can be thought of as belonging to one of five groups that differ dramatically when each group is compared to other groups. This means that ECS staff can grasp the essential diversity among science users by understanding the characteristics of only five groups rather than needing to understand hundreds or thousands of users. Readers can focus on the aspects of user diversity that are of particular interest to them by examining each group's mean response to questions that probe interesting issues. For example, a network designer might focus on each group's mean response to questions about whether or not the respondents expect to receive their data products by electronic network.

5.2 Basis for Founding Electronic Forums

As long as the ECS (or its successor systems) continues to serve Earth scientists we expect that the users to be at least as diverse as the results of the ESUS Project indicate. If ECS will always be serving diverse users, we should plan to proactively identify and accommodate that diversity. ESUS was designed as a way to initiate that process by empirically exploring user diversity. ESUS is but one preliminary survey that could be extended and updated by other surveys or be entirely replaced by fostering online dialog between the ECS and its users. If we choose online means to seek dialog with diverse users, we could invite hundreds or thousands of users to enter into dialog with ECS and its staff. But that option would require someone to synthesize those diverse and uncoordinated inputs into few enough messages that ECS can act upon those messages. As an alternative ECS could foster electronic forums that like-minded users can join and in which they can discuss ECS issues until approximate consensus emerges. Individuals who are uncomfortable with their forum's conclusions would be free to withdraw and join another more hospitable forum. After sufficient dialog has occurred among forum members, the forum would be encouraged to share the group's conclusions with ECS. Such user inputs would be reasonably well thought out and be shared by a group of a known size. It would behoove ECS to assimilate all such inputs from user forums.

If the foregoing theoretical proposal were implemented, ECS would have to select a mechanism for encouraging the formation of such forums. Two potential mechanisms follow: first, ECS could invite the most-typical members of each of the Groups 1-5 to chair forums that initially consist of the other members that were identified in Section 4.4.1. After those forums have been

organized, the members would be free to regroup if they found that their views were less similar than the ESUS analysis suggests. This approach has the advantage that the initial groups will reflect the best available knowledge about user diversity. Second, ECS could invite any scientist to volunteer a theme for and organize a forum. If many forums were proposed, the more popular forums would be allowed to continue after each has time to refine its unique views or agenda and recruit members. This approach could be successful if users feel motivated to found and join forums that represent all the main user viewpoints. But there is no guarantee that all main views will be spanned by the forums or that those views will be spanned as quickly as ECS might need to probe user views.

5.3 Use of Data for User Modeling

Some of the ESUS data is being factored into science user modeling in order to refine estimates of system performance. For example, data for questions 2.14 and 2.15 shows scientists expectations to receive their data via electronic networks or hard media, respectively. Those expectations imply pulls on ECS, which further imply system performance levels. Other questions that also may be used for user modeling include the system and data access questions (2.23-2.26).

6. Further Work

6.1 Extended ECS Observation of Typical Members of the Groups

As a paraphrase of one respondent's comment "surveys like this help but are not enough. You also need to observe scientists for 2 weeks." This scientist realized that every method of inquiry has its benefits and limits. Questionnaires, brief site visits, and extended site visits all have limitations, and all may be needed for different reasons. Before the ECS contract was awarded, various questionnaires were administered to scientists. After the ECS contract began, members of the Science Office and other staff have visited some three dozen scientists' research sites for a day or so. The suggestion for 2-week visits suggests that ECS might benefit from extended observations of some scientists by being able to acquire an in-depth understanding of how the scientists conduct their science and hearing concerns as they naturally surface. But it is not feasible for us to conduct extended observation of hundreds of scientists. We propose extended observation of the 2 most expert members of each group 1-5 totaling 10 site visits. Typical members of the identified groups might be sufficiently few scientists for ECS to consider this idea for extended observation of key scientists. These 10 scientists would collectively represent at least 140 of the 160 scientists and probably represent more scientists than 140.

In the worst case outcome, the ten scientists that are visited would be new randomly-chosen user consultants who, by chance, might provide valuable new insights about users. In the best case, the ten would emerge as the ten best representatives for ECS to consult whenever we want to understand the science user communities' views on issues such as definition of scenarios that describe their typical use of ECS. Whether the actual outcome is best, worst, or intermediate, the suggestion to make extended observation of ten "typical" scientists appears promising as a way to increase our sensitivity to diversity in user needs.

6.2 Expand the Present Data by Repeating the Survey with New Addresses

6.2.1 Send Questionnaire to Enlarged Population of Scientists

ECS might substantially enlarge its database of 160 scientist responses by using the 1994 EOS Directory. This increase would include scientists who were omitted from the previous directory and scientists who ESUS missed due to nondeliverable questionnaires.

The scientist population of respondents might be further enlarged by following one respondent's suggestion that the target population should not be limited to EOS Investigators. The scientist also elaborated his idea. "Why not point to it [the questionnaire] from the EOS [Mosaic] home page and announce it on the newsgroup sci.geo.eos?" This proposal potentially could enable ESUS to expand its data with responses from scientists whom we have not yet identified. By using the home page and a newsgroup we would be exploiting the Internet as a tool to prospect for as-yet-unidentified potential users.

We recommend that both means for obtaining new respondents be given serious consideration by ECS management.

6.2.2 Revise Questionnaire to Eliminate Language Problems

We recommend that the ESUS questionnaire should be slightly revised and administered following the proposals in Section 6.2. Those revisions would include the changes to which we committed in Section 4.2. The other alternatives are to cease surveying users and to design a new survey instrument after the survey goals and constraints have been identified. If there will be further surveying, we suggest the extremes of slight or complete revision would be feasible but an intermediate degree of revision would not be feasible. If we implement the fundamental criticisms in Section 4.2, we will need complete revision of the questionnaire and substantial time to complete that work. The value contained in this paper may demonstrate that further value can be obtained from slight revision and enlargement of the targeted scientist population.

6.2.3 Periodic Surveying of Scientists using the Revised Instrument

ECS might periodically probe its scientist users in order to map trends in scientists' attitudes about ECS beginning in 1994 when the scientists must imagine their uses of ECS because the system is not yet operational. Later when scientists know more about ECS and especially as they develop experience as users their attitudes will evolve. After scientists become experienced with Releases A, B, C, and D, we expect their attitudes about ECS to change. To catch such trends, ECS might administer some or many of the ESUS questions once or twice annually and reissue this Technical Paper after analyzing each repetition of the survey.

6.3 Rethink ECS's Strategy for Dialog

To date ECS has relied heavily on ECS visits of comparatively few scientists in order to dialog with our users. As ECS matures and the user population grows, we may become able to visit only a tiny fraction of the user population. ECS soon will become unable to rely on site visits because a few scientists are unlikely to be statistically representative of all scientists. Although site visits may always remain important to ECS, the time may be passing when we can rely on site visits as the main channel for dialog with users.

This paper demonstrates how users can be probed through a wider and more systematic sampling. User surveys clearly show promise as a way to supplement the site visit approach to obtaining dialog with users.

Section 5.2 proposed electronic forums as a means to facilitate dialogue among like-minded users as a preparation for their dialog with ECS concerning directions in which ECS should evolve. This approach is not systematic when compared to questionnaire surveying, but forums can involve large numbers of users unlike site visits.

ECS also could revise its user interfaces to allow users to comment about ECS or otherwise enter into online dialogue about ECS. The substantial volume of comments that ESUS respondents made suggests the possibility that users might actually welcome some opportunity to enter into dialog with ECS. The Release Plan presently includes "on-line collection of user feedback with

automatic prompting, combined with user satisfaction metrics for quantitative assessment." Evaluation Packages have been developed by ECS to allow user feedback on prototype versions of user-sensitive applications. These include an Interactive Evaluation Tool which allows users to enter comments at any time while using the system.

The foregoing paragraphs sketch some possible methods to achieve scientist-ECS dialog but do not determine the optimal mix of methods that should be implemented. We presently cannot identity that optimal mix and believe that the mix needs to be researched. We recommend that a study be conducted to take a broad look at ECS-User dialog and propose how to combine these and other methods. A well-integrated plan should be developed that will deliver the required level of dialog with an acceptable cost in ECS's system and human resources.

6.4 Further Analysis

In a low-level effort, continue the analysis of the existing ESUS data. For example, a more rigorous analysis should be made of the relationship between the respondents background traits (Questionnaire Section 1) and their membership in the clusters. This paper reports an ad hoc analysis that should be refined by a more rigorous statistical analysis perhaps via multiple discriminant analysis. Better knowledge of the relationship between background and attitudes about ECS could help us predict the data system attitudes of future science users. Also, somewhat more effective clusters analysis software has become available and should be used to check the clusters before ECS decides to settle on the exact list of scientists (by name) who would be invited to serve as representatives of each major group.

This page intentionally left blank.

Appendix A. E-Mail Version of Questionnaire

ECS Science User Survey

Introduction to User Survey

As a potential user of the EOSDIS (Earth Observing System Data and Information System) Core System (ECS), we urge you to participate in a survey that seeks feedback from the Earth science community on issues that will impact the ECS architecture.

You should be able to finish the survey within 30 minutes (slightly longer if you choose to contribute comments). The "questions" are actually requests for you to respond to "statements." After the last question you can request a survey report.

We would appreciate your responding by May 26, 1994 (but why not take it now?)

The following sections

- Explain the background of this survey.
- Encourage you to take the survey online using Mosaic but allow you to edit your responses into this message instead.
- Ask 15 questions that attempt to describe you as a user.
- Ask the remaining 26 questions that probe your views about the ECS and your potential use of ECS.
- Invite you to comment on the survey instrument.

If you need help, send E-mail to "eesus-help@eos.hitc.com" and we will respond.

Background of Survey

EOSDIS is a system whose design must be driven by user requirements. To that end, ECS must also be driven by those same requirements. However, user requirements can be elusive: user requirements change with the passage of time and accumulation of experience, articulation of requirements by users can be difficult, and users differ in their opinions. This survey is one of the instruments that ECS will be employing for discovery, refinement, and validation of users' requirements; its primary focus is on functional requirements that impact on the ECS architecture rather than on system performance.

Some surveys of the EOS community already have been conducted, and additional surveying is likely given the evolving nature of the community. We rely on such surveys to provide open channels through which users can communicate their needs.

This survey instrument was developed by Stan West, Mike Theobald, and Siri Jodha Singh Khalsa.

Instructions to Use Mosaic or E-mail

Mosaic is a graphical user interface with hypertext that is rapidly penetrating the Internet. We encourage users of all platforms that support Mosaic to use Mosaic Version 2 or later for this survey. After you take the survey we also will link you to several NASA information resources. (As of April 20, 1994 the Macintosh version of Mosaic is still version 1, so Macintosh users will need to take this survey by E-mail -- sorry.)

If you prefer to take the E-mail survey, just edit this message and send to:

eesus@eos.hitc.com.

If you already use the latest UNIX or Microsoft Windows version of Mosaic and want to take the Mosaic version of the survey, you should:

Run Mosaic

Click on the "File" pulldown menu

Select "Open URL..."

Type "http://boreas.colorado.edu/eos.survey.html" in the URL box

Click "Ok"

If you need to install Mosaic, follow these steps for ftp and installation:

"ftp ftp.ncsa.uiuc.edu"

login as "anonymous" (line occasionally may be busy)

type your email address as password

"cd Mosaic"

"cd Windows" for PC, "cd Mosaic-binaries" for UNIX, or "cd Mac" for

Macintosh (only use Macintosh if Version 2 has been released)

Follow the README instructions to "get" the binary file for Mosaic

Install Mosaic following the directions provided with Mosaic

Ignore the rest of this message if you plan to take the survey with Mosaic.

Section 1. Questions about Your Background

Please type your name and E-mail address if you do NOT plan to use Mosaic.

Name: _____

E-mail address (if different than address in message header):

_____

Please check (X) all of the statements in Section 1 that pertain to you.

Where appropriate, check more than one statement in any of the following

groups of statements.

1.1 ( ) I am an instrument developer.

1.2 ( ) I am a data product developer.

1.3 ( ) I am a data product consumer.

1.4 ( ) The geographic scale of my primary research interest is local.

1.5 ( ) The geographic scale of my primary research interest is regional.

1.6 ( ) The geographic scale of my primary research interest is global.

1.7 ( ) The time scale of my primary research interest is decades.

1.8 ( ) The time scale of my primary research interest is years.

1.9 ( ) The time scale of my primary research interest is months.

1.10 ( ) The time scale of my primary research interest is days.

EOS scientists are likely to develop tools for scientific analysis and could share those tools with other scientists.

1.11 ( ) I plan to contribute a science analysis tool to ECS.

1.12 ( ) I would use science analysis tools that other scientists might contribute to ECS.

The number of data sets that I work with at one time is usually

1.13 ( ) One

1.14 ( ) A few

1.15 ( ) Many

Section 2. Questions Probing Your Views about ECS

In this section we seek your advice about the capabilities that ECS should provide to science users. The following statements support some capabilities that are already-planned and other capabilities that are not yet planned.

Please answer to the best of your ability, but if you have difficulty with any question, help us understand why by contributing a comment at the bottom of the message.

Use this scale to agree or disagree with each statement in this section:

Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
5	4	3	2	1

2.1 ( ) ECS should support representation of Level 3 and Level 4 data products on a standard set of grids.

2.2 ( ) ECS should support representation of Level 3 and Level 4 data products on user-specifiable grids.

2.3 ( ) ECS data products should have one standard data format for data distribution.

2.4 ( ) ECS should have multiple standard data formats for data distribution.

2.5 ( ) ECS should support user-specifiable formats for data distribution.

2.6 ( ) ECS should provide software for translation between data formats.

2.7 ( ) ECS should provide access to data developed at Science Computing Facilities.

2.8 ( ) I would like ECS to facilitate other scientists' access to my data.

2.9 ( ) ECS should incorporate scientist-contributed tools.

2.10 ( ) ECS should provide some tools for data visualization.

2.11 ( ) ECS should provide capability for video conferencing.

2.12 ( ) ECS should provide capability for audio annotation of data.

2.13 ( ) ECS should provide capability for video annotation of data.

2.14 () I expect most ECS data to be delivered to me via electronic networks.

2.15 ( ) I expect most ECS data to be delivered to me via tapes, CD-ROMs, etc.

The following statements explore tradeoffs between data product quality and other goals, such as timeliness of availability. "Quality" is a broad term and includes measures of accuracy and the use of algorithms that contain newer, but perhaps still experimental, treatments of physical processes.

2.16 ( ) I would be willing to wait 1 hour for my data to be produced after specifying what I want.

2.17 ( ) I would be willing to wait 1 day for my data to be produced after specifying what I want.

2.18 ( ) I would be willing to wait 1 week for my data to be produced after specifying what I want.

2.19 ( ) I need data products that use the latest updates in science algorithms.

2.20 ( ) I need data products that have long-term consistency.

2.21 ( ) I need data products as soon as they are available.

2.22 ( ) I need data products of verified high quality.

The following two statements explore different models of system access. In the first model, users access data by placing an order for it and then having that order satisfied via electronic or physical media. In the second, users access data without submitting an order.

2.23 ( ) I would prefer to place orders for delivery of ECS data to me.

2.24 ( ) I would prefer to have automatic access to ECS data after I specify what I want.

The following two statements also deal with data access issues. The first statement suggests that ECS should provide resources for additional data analysis before those data are delivered to a user. The second statement investigates interest in the ability to place standing orders for data in advance of data production.

2.25 ( ) I expect to perform operations on data without first having them transferred to me.

2.26 ( ) I expect to submit standing orders for data so that as they are collected and produced they will be delivered to me.

( ) Check (X) if you would like to receive a report describing the results of this survey.

Please use the space below to elaborate on any strong feelings that you have about any particular statements. Identify the statement numbers when you comment. Also feel free to comment on any additional issues that you think should be addressed in this questionnaire.

(Type as many lines as you like beginning on the next line)

(End of Comments)

This page intentionally left blank.

Abbreviations and Acronyms

ASCII	American Standard Code for Information Interchange
CDROM	Compact Disk, Read Only Memory
CPU	Central Processing Unit
DAAC	Distributed Active Archive Center
ECS	EOSDIS Core System
EOS	Earth Observing System
EOSDIS	EOS Data Information System
ESUS	ECS Scientist User Survey
ftp	File Transfer Protocol
HDF	Hierarchical Data Format
IEEE	Institute of Electrical and Electronic Engineers
I/O	Input/Output
NASA	National Aeronautics and Space Administration
SCF	Science Computing Facilities
SDR	System Design Review
SRR	System Requirements Review
UARS	Upper Atmospheric Research Satellite
UK	United Kingdom
URL	Universal Resource Locator